

УДК 004.652.3

**МЕТОДЫ ИССЛЕДОВАНИЯ СТРУКТУРЫ
МЕДИЦИНСКИХ ДАННЫХ**

О.Г. Берестнева, И.А. Осадчая, Е.В. Немеров*

Томский политехнический университет
*Сибирский государственный медицинский
университет, г. Томск
E-mail: ogb6@yandex.ru

Рассмотрены методы структурного анализа многомерных данных (кластерный анализ и методы визуализации) в медицине. Проанализированы особенности применения различных методов исследования структуры экспериментальных данных в медицине. Представлены результаты применения методов визуализации для выявления особенностей различных форм бронхиальной астмы.

Ключевые слова:

Кластерный анализ, когнитивная графика, структура медицинских данных.

Введение

В настоящее время накоплен обширный арсенал средств анализа многомерных данных. Наиболее полное изложение применяемых здесь подходов, сопровождающееся подробными ссылками на ключевые работы, содержится в [1]. В [2] приведена классификация основных методов анализа структуры многомерных данных.

Выделяют:

1. Методы визуализации данных:
 - линейные методы снижения размерности;
 - нелинейные отображения;
 - многомерное шкалирование;
 - заполняющие пространство кривые.
2. Методы автоматического группирования:
 - факторный анализ объектов и признаков;
 - кластерный анализ объектов и признаков;

- иерархическое группирование;
- определение «точек сгущения».

Разделение методов носит достаточно условный характер, так как различные методы имеют немало пересечений в отдельных приемах обработки информации. В основу приведенной классификации положен признак, отображающий степень участия экспериментатора в выделении особенностей взаимоотношений между исследуемыми объектами и признаками. Применение методов визуализации данных нацелено на поиск наиболее выразительных изображений совокупности исследуемых объектов для последующего максимального задействования потенциала зрительного анализатора экспериментатора.

Настоящая работа посвящена вопросам применения методов структурного анализа данных в медицине. В работе рассмотрены возможности методов когнитивной графики и кластер-

Берестнева Ольга Григорьевна, д-р техн. наук, профессор кафедры прикладной математики Института кибернетики ТПУ, профессор кафедры общей и дифференциальной психологии Сибирского государственного медицинского университета, г. Томск.
E-mail: ogb6@yandex.ru

Область научных интересов: математическое моделирование, медицинские информационные технологии, когнитивная графика.

Осадчая Ирина Александровна, студентка кафедры прикладной математики Института кибернетики ТПУ.
E-mail: irishka_tomsk@mail.ru

Область научных интересов: математическое моделирование, медицинские информационные технологии, когнитивная графика.

Немеров Евгений Владимирович, канд. мед. наук, ассистент кафедры поликлинической терапии Сибирского государственного медицинского университета, г. Томск.
E-mail: nemerevg@mail.ru

Область научных интересов: поликлиническая терапия, медицинские технологии.

ного анализа для исследования структуры медицинских данных (на примере задачи исследования особенностей различных форм бронхиальной астмы).

Характеристика экспериментальных данных

Проведем анализ показателей физиологических реакций бронхолегочной системы в ответ на психофизиологическое воздействие (аудиовизуальную стимуляцию). Технология получения экспериментальных данных подробно изложена в [3].

Бронхиальная астма (греч. *asthma* – удушье) – болезнь, при которой у больных наблюдаются приступы экспираторного удушья различной тяжести [4]. Бронхиальная астма может возникать под воздействием целого ряда внешних и внутренних причин.

Большинство больных бронхиальной астмой страдают психогенными приступами удушья. Такие приступы возникают, как правило, при сильных эмоциях страха или гнева. Тяжелые стрессовые ситуации, скорее всего, могут вызвать непродолжительную ремиссию бронхиальной астмы. В свою очередь, хронические психотравмы в большинстве случаев попросту ухудшают её течение [4].

Одной из причин возникновения различных соматических заболеваний могут быть психические факторы. Важной группой таких факторов являются негативные эмоции. Эмоции оказывают влияние на: иммунную систему; гормональное состояние; периферическую физиологическую активацию (например, частоту сердцебиений и артериальное давление) [5].

Исходная информация представляет собой данные о пациентах с четырьмя типами бронхолегочных заболеваний:

- Бронхиальная астма непсихогенная (*BANP*)
- Бронхиальная астма сомато-психогенная (*BASP*)
- Бронхиальная астма психогенно индуцированная (*BAPI*)
- Психогенная одышка (*PD*).

Анализ данных с помощью методов когнитивной графики

Традиционные инструменты в области визуализации (графики и диаграммы) плохо справляются со своей задачей, когда возникает необходимость изобразить более трех взаимосвязанных величин. Методы многомерного анализа – наиболее действенный количественный инструмент исследования социально-экономических процессов, описываемых большим числом характеристик [6]. К ним относятся кластерный анализ, таксономия, распознавание образов, факторный анализ.

Компьютерная обработка данных предполагает некоторое математическое преобразование данных с помощью определенных программных средств. Для этого необходимо иметь представление как о математических методах обработки данных, так и о соответствующих программных средствах [7].

На данный момент существует множество различных пакетов, предназначенных для анализа экспериментальных данных в различных областях человеческой деятельности. Наиболее известными являются компьютерные системы SAS, SPSS, SYSTAT, Minitab, Statgraphics (зарубежные). В России популярны пакеты Statistica/W и STADIA.

Использование компьютера для обработки данных делает достаточно сложные методы анализа данных более доступными и наглядными. Поэтому исследователю остается в основном творческая работа: постановка задач, выбор методов их решения, интерпретация полученных результатов, выработка предложений и управленческих решений.

Применение графики в исследовательских работах не только увеличивает скорость передачи информации и повышает уровень ее понимания, но и способствует развитию таких важных для специалиста любой отрасли качеств, как интуиция, образное мышление.

Воздействие интерактивной компьютерной графики (ИКГ) привело к возникновению нового направления в проблематике искусственного интеллекта, названного когнитивной (т. е. способствующей познанию) компьютерной графикой.

Когнитивная графика – это совокупность приемов и методов образного представления условий задачи, которое позволяет либо сразу увидеть решение, либо получить подсказку для его нахождения [6].

Использование когнитивной графики дает возможность пользователю, не анализируя большого количества информации, сделать определенные выводы. Информация может быть представлена когнитивным образом: сектором, гистограммой, крестом, кругом и т. д., части которых закрашены разными цветами и несут определенный смысл.

Отдельное направление когнитивная графика образует в медицине. Визуализация текущего состояния объекта и характерных особенностей позволяет обеспечить непрерывный контроль над состоянием групп лиц либо отдельного человека.

Нами был использован подход, предложенный В.А. Воловоденко [8], позволяющий отображать многомерные объекты в виде кривых или «спектров». «Спектральные представления» в данном методе подчеркивают отличительные характеристики каждой кривой и помогают более детально исследовать их визуальные свойства. Цветная палитра акцентирует уровни изменения значений кривых. Производя воображаемое растяжение кривых вдоль Z-оси и глядя сверху на результат этой операции, можно получить цветные полосы, представляющие собой спектр каждого наблюдения [8].

Сравним выборочные данные для четырех форм бронхиальной астмы по визуальной близости спектров наблюдений (рис. 1). Результаты получены в пакете «NovoSparkVisualizer».

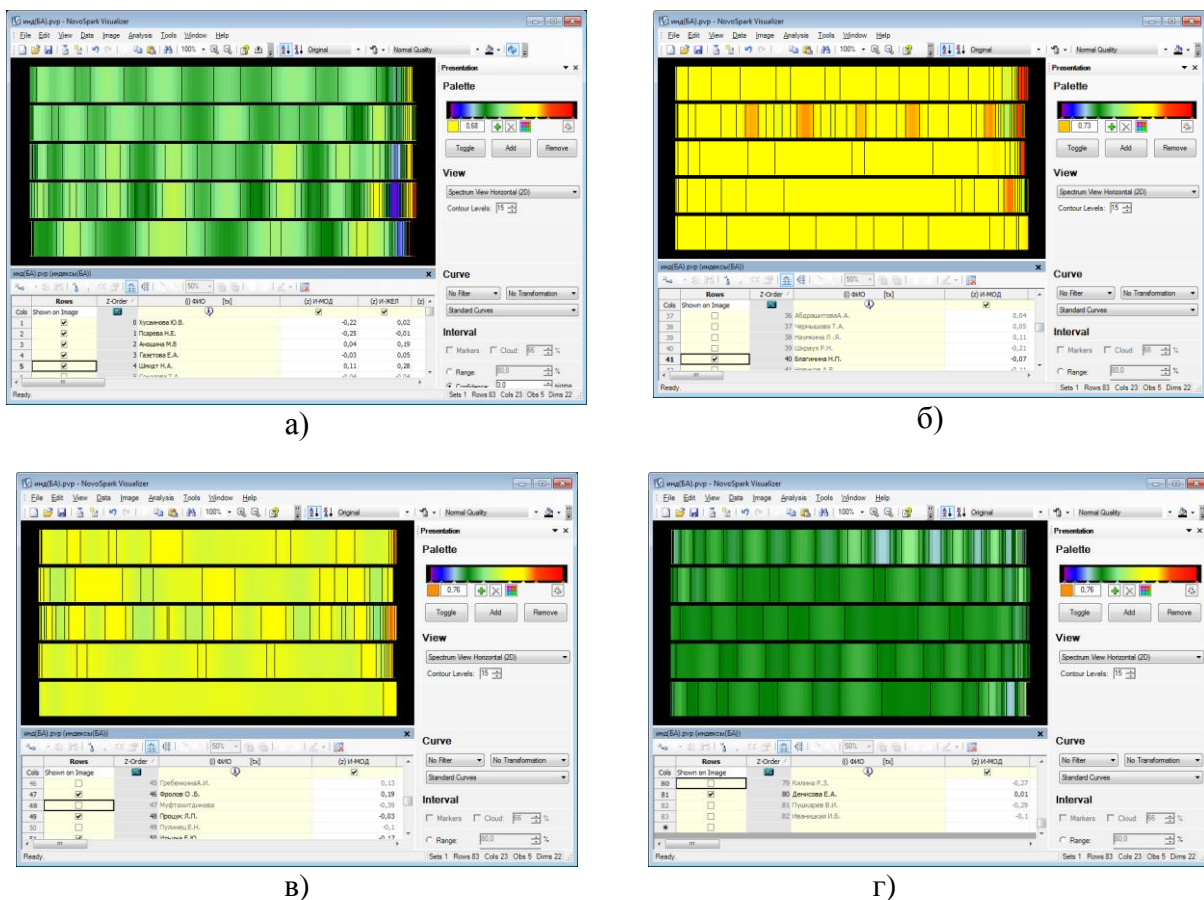


Рис. 1. Спектральные представления данных о пациентах с различными диагнозами: а) *BAPI*; б) *BASP*; в) *BANP*; г) *PD*

Каждая цветная полоска в спектральном виде соответствует показателям одного пациента. В нашем случае на рис. 1 представлено по пять цветных полосок, соответственно представляющих по пять пациентов с различными формами бронхиальной астмы. Цветовые «спектры» пациентов с диагнозом *BAP1* и *PD* схожи. Тоже можно сказать и о пациентах с диагнозом *BASP* и *BANP*.

Как видно из рисунков, наиболее близкие показатели имеют пациенты с *BASP* и *PD*. Наиболее выраженные различия можно заметить у пациентов с *BAP1*.

Таким образом, использование средств когнитивной графики позволило выявить некоторые не известные ранее закономерности физиологических реакций бронхолегочной системы в ответ на психофизиологическое воздействие.

Анализ данных на основе кластерного анализа

Кластер (англ. *cluster* – скопление) – объединение нескольких однородных элементов, которое может рассматриваться как самостоятельная единица, обладающая определёнными свойствами. Кластерный анализ наиболее ярко отражает черты многомерного анализа в классификации. Главное назначение кластерного анализа – разбиение множества исследуемых объектов и признаков на однородные, в соответствующем понимании, группы или кластеры. Это означает, что решается задача классификации данных и выявления соответствующей структуры в ней. Методы кластерного анализа можно применять в самых различных случаях, даже в тех случаях, когда речь идет о простой группировке, в которой все сводится к образованию групп по количественному сходству [2]. Основное достоинство кластерного анализа в том, что он позволяет производить разбиение объектов не по одному параметру, а по целому набору признаков. В табл. 1 и 2 представлены результаты разбиения, полученные с помощью кластерного анализа (метод Уорда) на базе пакета *Statistica*.

В табл. 1 представлены результаты кластеризации по физиологическим показателям бронхолегочной системы, отражающим динамику их изменения после проведения сеансов аудиовизуальной стимуляции.

Таблица 1. Распределение пациентов по кластерам (степень воздействия аудиовизуальной стимуляции)

Диагноз	Количество пациентов			
	1 кластер	2 кластер	3 кластер	4 кластер
BANP	1	5	15	8
BASP	3	6	5	4
BAP1	8	14	2	0
PD	1	0	6	5

В 1-м кластере преимущественно оказались пациенты с диагнозом психогенно индуцированной бронхиальной астмой (ВА), а также 3 человека с сомато-психогенной ВА, 1 – с непсихогенной ВА и 1 – с психогенной отдышкой.

Во второй кластер преимущественно попали также больные психогенно индуцированной бронхиальной астмой, 6 человек – с диагнозом сомато-психогенной бронхиальной астмы и 5 человек – с непсихогенной астмой.

Третий кластер объединил преимущественно людей с диагнозом непсихогенной бронхиальной астмы. Также сюда попали 2 человека с психогенно индуцированной астмой, 5 человек – с сомато-психогенной астмой и 6 – с психогенной отдышкой.

В четвертом кластере также преимущественно оказались люди с непсихогенной бронхиальной астмой. А также 4 – с сомато-психогенной ВА и 5 – с психогенной отдышкой.

На основе анализа полученных результатов (значений центроидов для каждого кластера) был сделан вывод о том, что на пациентов 1, 3 и 4 кластеров аудиовизуальная стимуляция мозга оказывает аналогичное воздействие – улучшаются значения показателей вентиляции лег-

ких и механики дыхания. При этом для первого кластера улучшение значений показателей вентиляции легких менее выражено, чем для третьего кластера. Улучшение статической растяжимости легких характерно только для 4 кластера. Для пациентов, попавших во 2 кластер, аудиовизуальная стимуляция практически не оказывает никакого воздействия на изменение показателей вентиляции легких и механики легких.

В табл. 2 представлены результаты кластеризации по исходным («фоновым») физиологическим показателям бронхолегочной системы.

Таблица 2. Распределение пациентов по кластерам («фоновые» показатели бронхолегочной системы)

Диагноз	Количество пациентов			
	1 кластер	2 кластер	3 кластер	4 кластер
BANP	6	21	0	2
BASP	3	6	7	2
BAP1	14	1	7	2
PD	1	8	2	1

В первый кластер попали, в основном, пациенты с диагнозом «психогенно индуцированная астма», во второй – пациенты с непсихогенной бронхиальной астмой и с психогенной одышкой, в третий – пациенты с психогенно-индуцированной и сомато-психогенной бронхиальной астмой. И, наконец, в четвертом оказалось практически одинаковое количество представителей каждого из заболеваний. Полученные результаты показывают, что на основе имеющихся показателей бронхолегочной системы не удалось выделить кластеры, соответствующие медицинским диагнозам.

Выводы

На основе проведенного структурного анализа медицинских данных (когнитивная графика и кластерный анализ) можно сделать следующие выводы:

1. С точки зрения структуры показателей бронхолегочной системы можно выделить две однородные группы пациентов:
 - с диагнозами *BAP1* и *PD*
 - с диагнозами *BANP* и *BASP*
2. Применение аудиовизуальной стимуляции мозга наименее эффективно для пациентов с диагнозами *BANP* и *PD*, наиболее эффективно – для пациентов с диагнозом *BAP1*. Для пациентов с диагнозом *BASP* общей тенденции не выявлено.

СПИСОК ЛИТЕРАТУРЫ

1. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Классификация и снижение размерности. – М.: Финансы и статистика, 1989. – 471 с.
2. Дюк В., Эмануэль В. Информационные технологии в медико-биологических исследованиях. – СПб.: Питер, 2003. – 528 с.
3. Немеров Е.В., Языков К.Г. К вопросу изучения личностных свойств в психофизиологической реактивности больных бронхиальной астмой на аудиовизуальную стимуляцию // Вестник ТГПУ. – 2011. – Вып. 6 (108). – С. 134–137.
4. Бронхиальная астма // Бронхиальная астма // Candidoz.Ru (молочница). 2011. URL: <http://www.candidoz.ru/allergologiya/astma.html> (дата обращения: 06.01.2012).
5. Медико-психологические аспекты психосоматических соотношений в исследованиях отечественных авторов последнего периода // Психология человека. 2012. URL: <http://psibook.com/17/20.html> (дата обращения: 15.01.2012).

6. Зенкин А.А. Когнитивная компьютерная графика / под ред. Д.А. Поспелова. – М.: Наука, 1991. – 187 с.
7. Тарасов В.Б. От многоагентных систем к интеллектуальным организациям: философия, психология, информатика. –М.: Эдиториал УРСС, 2002. – 352 с.
8. Воловоденко В.А. Визуализация и анализ многомерных данных с использованием пакета «NovoSparkVisualizer» // www.novospark.com. 2008. URL: <http://www.tsu.ru/storage/iro/k020410/s4/s4.doc> (дата обращения 25.11.2011).
9. Шаропин К.А., Берестнева О.Г., Воловоденко В.А., Марухина О.В. Визуализация медицинских данных на базе пакета NovoSpark // Известия Южного федерального университета. Технические науки. – 2010. – Т. 109. – № 8. – С. 242–249.

Поступила 20.01.2012 г.