

УДК 004.92

МЕТОДЫ ИССЛЕДОВАНИЯ СТРУКТУРЫ МНОГОМЕРНЫХ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Осадчая И. А., Берестнева О. Г.

*ФГБОУ ВПО Национальный исследовательский Томский политехнический университет, Томск, Россия
(634050, Томск, проспект Ленина, 30), e-mail: ogb@tpu.ru*

При анализе данных исследователь довольно часто сталкивается с многомерностью их описания. Методы многомерного анализа – наиболее действенный количественный инструмент исследования процессов, описываемых большим числом характеристик. В работе рассмотрены особенности структурного анализа данных с использованием кластерного анализа и когнитивной графики. Приведены примеры решения двух прикладных задач: структурный анализ показателей механики дыхания у больных с различными типами бронхиальной астмы и показателей мотивационной сферы студентов гуманитарных и технических специальностей. Для компьютерной обработки и анализа данных использовались прикладные пакеты Statistica и NovoSpark Visualizer. Результаты проведенных авторами исследований позволяют утверждать, что представленные в статье методы и подходы являются перспективными направлениями в области анализа и представления многомерных экспериментальных социально-психологических и медицинских данных.

Ключевые слова: методы визуализации, когнитивная графика, кластерный анализ, структура медицинских данных.

METHODS FOR STUDYING THE MULTIDIMENSIONAL EXPERIMENTAL DATA STRUCTURE

Osadchaya I. A., Berestneva O. G.

*National Research Tomsk Polytechnic University, Tomsk, Russia (634050, Tomsk, avenue of Lenina, 30), e-mail:
ogb@tpu.ru*

In analyzing the data the researcher often faced with multi-dimensional descriptions. Multivariate analysis - the most effective quantitative research tool processes described by a large number of characteristics. This paper describes the features of the structural analysis of the data using cluster analysis and cognitive graphics. Examples of solutions of two applications: a structural analysis of indicators of respiratory mechanics in patients with different types of asthma and indicators of students' motivational sphere of humanitarian and technical specialties. For computer processing and data analysis software packages used Statistica and NovoSpark Visualizer. The results of the study authors suggest that the methods presented in the paper and approaches are promising directions in the field of multi-dimensional analysis and experimental psycho-social and medical data.

Keywords: imaging, cognitive graphics, cluster analysis, medical data structure.

При анализе и прогнозировании различных явлений исследователь довольно часто сталкивается с многомерностью их описания. Методы многомерного анализа – наиболее действенный количественный инструмент исследования процессов, описываемых большим числом характеристик [1, 4, 8].

Выделяют две основные группы методов анализа структуры многомерных данных [5]:

- 1) визуализация данных: линейные методы снижения размерности, нелинейные отображения, многомерное шкалирование, заполняющие пространство кривые;
- 2) автоматическое группирование: факторный и кластерный анализ объектов и признаков, иерархическое группирование, определение «точек сгущения».

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.

Наиболее часто структурный анализ многомерных экспериментальных данных применяется в слабоструктурированных проблемных областях (социальная сфера, психология, медицина).

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 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.

Основной задачей визуализации данных является задача получения визуального образа, однозначно соответствующего набору данных. Нами был использован подход, предложенный В. А. Воловоденко [1, 4, 8], позволяющий отображать многомерные объекты в виде кривых или «спектров». Применение предложенного метода, на наш взгляд, приводит к значительному упрощению процедуры визуализации и способствует выявлению скрытых связей в очевидной форме. Это позволяет не только использовать свойства метода, но и способствует учету опыта группы экспертов, рассматривающих конкретную проблему. Последнее обстоятельство повышает актуальность предложенного подхода.

«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.

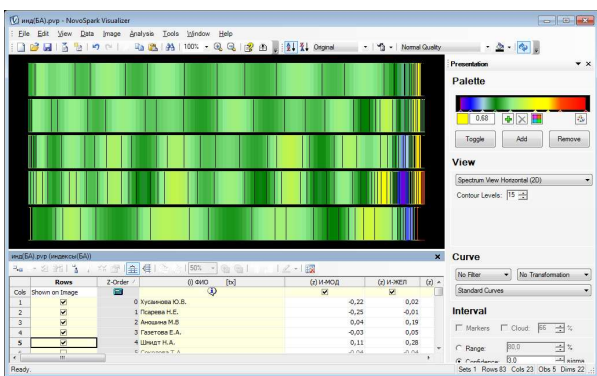
Рассмотрим особенности решения задачи структурного анализа на примере экспериментальных медицинских и психологических данных.

В качестве исходной информации в первом случае имеем данные о динамике показателей вентиляции легких и механики дыхания в ответ на психофизиологическое воздействие (аудиовизуальную стимуляцию) [7].

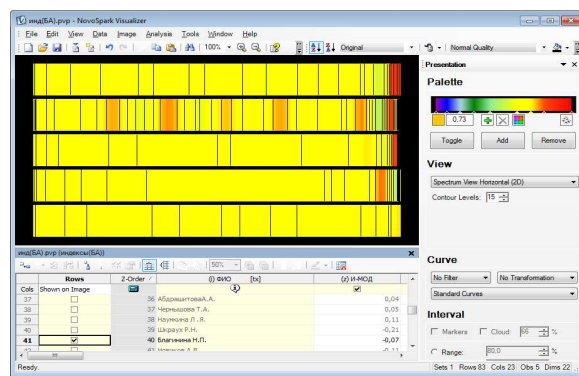
Таким образом, исходная информация представляет собой данные о пациентах с четырьмя типами бронхолегочных заболеваний:

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

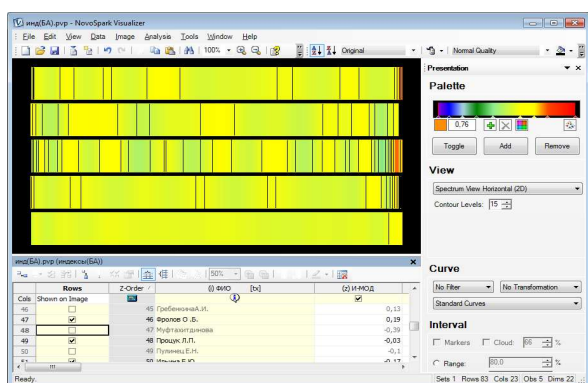
Сравним выборочные данные для четырех форм бронхиальной астмы по визуальной близости спектров наблюдений (рис.1). Результаты получены в пакете «*NovoSparkVisualizer*».



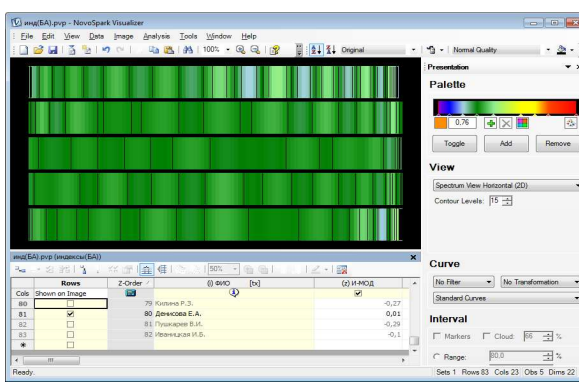
а)



б)



с)



д)

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

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

Наиболее ярко отражает черты многомерного анализа в классификации кластерный анализ. Кластер – объединение нескольких однородных элементов, которое может рассматриваться как самостоятельная единица, обладающая определёнными свойствами. 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 [8]. 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

Диагноз	Количество пациентов			
	1 кластер	2 кластер	3 кластер	4 кластер
<i>BANP</i>	1	5	15	8
<i>BASP</i>	3	6	5	4
<i>BAPI</i>	8	14	2	0
<i>PD</i>	1	0	6	5

В 1-м кластере преимущественно оказались пациенты с диагнозом *BAPI*, а также 3 человека с *BASP*, 1 – с *BANP* и 1 – с *PD*. Во второй кластер преимущественно попали также больные *BAPI*, 6 человек – с диагнозом *BASP* и 5 человек – с *PD*. Третий кластер объединил преимущественно людей с диагнозом *BANP*. Также сюда попали 2 человека с *BAPI*, 5

человек – с *BASP* и 6 – с *PD*. В четвертом кластере также преимущественно оказались люди с *BANP*, а также 4 – с *BASP* и 5 – с *PD*.

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 легких 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. У 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 показателям механики дыхания.

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

Диагноз	Количество пациентов			
	1 кластер	2 кластер	3 кластер	4 кластер
<i>BANP</i>	6	21	0	2
<i>BASP</i>	3	6	7	2
<i>BAP1</i>	14	1	7	2
<i>PD</i>	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 отдышкой, 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.

Рассмотрим возможности применения кластерного анализа на примере анализа структуры экспериментальных данных, полученных в результате исследования мотивационной сферы учебной деятельности студентов.

В структуре личности мотивация – основное понятие, используемое для объяснения движущихся сил поведения и деятельности человека. Мотив всегда существует в деятельности человека до тех пор, пока не достигнута цель, либо изменившиеся условия не сделают другой мотив более насущным для данного человека. Для более полной оценки перспектив общего мотивационного развития в процессе обучения или профессионального становления исследуют характер соотношения основных мотивационных тенденций. В нашем случае была использована методика «Мотивационный профиль личности» [2], позволяющая диагностировать

следующие мотивы учебной деятельности: П – поддержание жизнеобеспечения; К – комфорт; С – социальный статус; О – общение; Д – общая активность; ДР – творческая активность; ОД – общественная полезность.

По данной методике были обследованы 3 группы студентов:

- гуманитарного факультета Томского политехнического университета;
- бизнес-инкубатора Томского политехнического университета;
- психологического факультета Томского государственного педагогического университета.

В результате кластерного анализа методом k-средних было выделено 4 кластера, характеристики которых представлены в таблице 3.

Таблица 3. Центроиды кластеров

Кластер	П	К	С	О	Д	ДР	ОД
1	10,5	14,1	24,3	21,3	7,5	17,7	-0,75
2	17	20,4	32,8	28	17,5	23,5	13,8
3	3,1	4,2	11,0	11,7	-0,8	8,5	-5,4
4	25	41	55	42	30	38,5	25,5

Очевидно, что людей, находящихся в том или ином кластере, объединяют отличительные характеристики. Первый кластер объединил студентов, имеющих высокие мотивы общения и социального статуса, а также довольно низкий показатель общественной полезности. В эту группу вошли студенты гуманитарного факультета и студенты-психологи.

Второй кластер объединил творчески активных студентов, имеющих высокие показатели мотива поддержания жизнеобеспечения и социального статуса, а также студентов, стремящихся к общению. В эту группу вошли студенты бизнес-инкубатора ТПУ. Третий кластер объединил студентов, имеющих самые низкие показатели общей активности и общественной полезности. В данную группу попали представители каждой из групп. В четвертом кластере оказался студент из первой группы, имеющий высокие показатели по каждой характеристике.

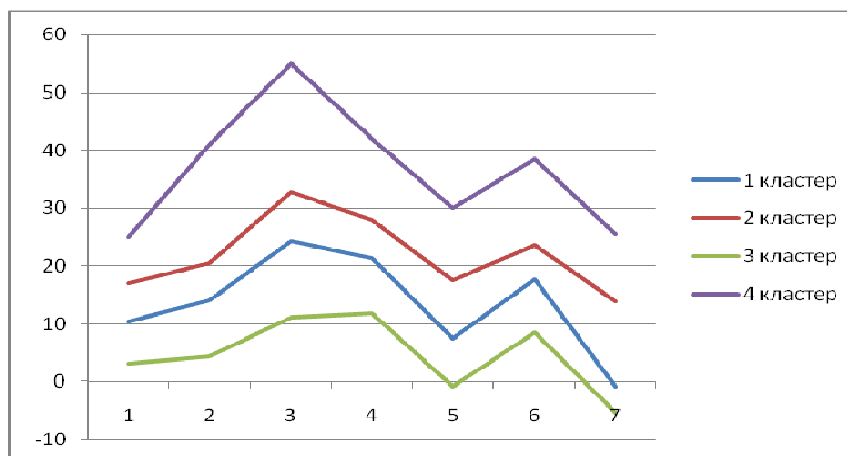


Рис. 2. График центроидов кластеров

Таблица 4. Количественное присутствие студентов группы в кластере

	1кластер	2кластер	3кластер	4кластер
1группа	13	8	6	1
2группа	6	1	9	0
3группа	5	0	11	0
4группа	7	2	5	0

Студенты 2 и 3 групп попали преимущественно в 3 кластер, т.е. для них характерны отрицательные значения показателей общественной активности (Д) и общественной полезности (ОД), т.е. неспособность к планомерной учебной деятельности и нежелание принимать участие в общественной жизни.

Студенты 1 и 4 групп оказались преимущественно в 1-м кластере, что говорит о большом стремлении к общению и желании повысить свой социальный статус.

В таблицах представлены данные о структуре полученных кластеров. Анализ полученных результатов показал, что в первый кластер попали студенты из всех групп, т.е. его характеристики можно условно считать типичным мотивационным профилем для студентов 3 – 5 курсов.

В соответствии с методикой «Мотивационный профиль личности» выделяется 4 типа мотивационных профилей [2]:

1 – прогрессивный

3 – импульсивный

2 – регрессивный

4 – уплощенный

Как видно из графика центроидов кластеров, в нашем случае наблюдается только импульсивный тип мотивационного профиля. При этом структура мотивов во всех кластерах одинакова – наиболее выраженными являются мотивы С и О, наименее выраженными – мотивы Д и ОД.

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

1. Спектральное представление визуального образа является более «тонким» инструментом, подчеркивающим различия или сходства образов, чем традиционные методы, характеризующие эти свойства на уровне числовых параметров. Общие цветовые различия можно игнорировать путем перехода к монохроматическому представлению.

2. С точки зрения структуры показателей бронхолегочной системы можно выделить две однородные группы пациентов:

- с диагнозами *BAP1* и *PD*;
- с диагнозами *BANP* и *BASP*.

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.

3. Кластерный анализ мотивационной сферы личности студентов выявил импульсивный тип мотивационного профиля.

Результаты проведенных исследований позволяют утверждать, что представленные в статье методы и подходы являются перспективными направлениями в области анализа и представления многомерных экспериментальных данных.

Список литературы

1. Берестнева О. Г., Воловоденко В. А. Визуализация экспериментальных многомерных данных на основе обобщенных графических образов // Вестник науки Сибири. – 2011. – № 1 (1). – С. 363-369.
2. Берестнева О. Г., Дубинина И. А., Уразаев А. М. Психологическое тестирование. – Томск: Изд-во Томского политехнического университета, 2005. – 144 с.
3. Берестнева О. Г., Осадчая И. А., Немеров Е. В. Методы исследования структуры медицинских данных // Вестник науки Сибири. – 2012. – № 1 (2). – С. 333-338.
4. Воловоденко В. А., Берестнева О. Г. Методы визуализации в оценке состояния человека-оператора // Известия Южного федерального университета. Технические науки. – 2012. – Т. 134. – № 9. – С. 9–13.
5. Дюк В., Эмануэль В. Информационные технологии в медико-биологических исследованиях. – СПб.: Питер, 2003. – 528 с.
6. Зенкин А. А. Когнитивная компьютерная графика / Под ред. Д. А. Поспелова. – М.: Наука, 1991. – 187 с.
7. 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.
8. 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.

Рецензенты:

Иванкина Любовь Ивановна, д.ф.н., профессор, Национальный исследовательский Томский политехнический университет, г. Томск.

Коваль Тамара Васильевна, д.ф.-м.н., профессор, Институт кибернетики Национального исследовательского Томского политехнического университета, г. Томск.