

УДК 004.9

ВИЗУАЛИЗАЦИЯ ЭКСПЕРИМЕНТАЛЬНЫХ МНОГОМЕРНЫХ ДАнных НА ОСНОВЕ ОБОБЩЕННЫХ ГРАФИЧЕСКИХ ОБРАЗОВ

О.Г. Берестнева, В.А. Воловоденко, К.А. Шаропин,
О.М. Гергет

Томский политехнический университет
E-mail: ogb@tpu.ru

Представлены различные подходы к визуализации результатов экспериментальных исследований. Приведены примеры решения прикладных задач с использованием *NovoSpark Visualizer*.

Ключевые слова:

Методы визуализации, многомерные экспериментальные данные.

Key words:

Imaging, multi-dimensional experimental data.

Современные компьютерные технологии используют широкий спектр методов визуализации информации. Легкость построения графиков и диаграмм с помощью ЭВМ все заметнее меняет когнитивные навыки исследователя. Современные пакеты анализа информации позволяют строить сотни типов различных графиков и диаграмм, в том числе так называемые пиктографики «Лица Чернова». Результаты применения последних в анализе психофизиологических данных представлены в [1]. Методы визуализации позволяют исследователю одним взглядом обнаружить особенности, выявить закономерности и аномалии в больших объемах информации.

Особенно широко графические методы используются в разведочном анализе данных и задачах кластеризации, позволяя выявлять закономерности в многомерных массивах информации. Методы

визуализации такого рода (метод главных компонент [2], методы структурного упорядочения [3, 4] и др.) основаны, как правило, на переходе от многомерной к двумерной системе координат.

Основной задачей визуализации данных является задача получения визуального образа, однозначно соответствующего набору данных. Например, в методе структурного упорядочения таким визуальным образом является точка в преобразованном пространстве координат. В данной статье рассмотрен подход, предложенный В.А. Воловоденко [4, 5], который позволяет провести визуализацию основных линейных конструкций: отрезок, ломанная линия, симплекс в многомерных пространствах. Основой визуализационного подхода является линейное преобразование значений многомерного наблюдения A в двумерную кривую $f_A(t)$, т. е. $A \leftrightarrow f_A(t)$, при этом гарантируется, что близким по значениям наблюдениям A и B будут соответствовать визуально близкие образы-кривые $f_A(t)$ и $f_B(t)$; для сильно различающихся по значениям наблюдений их образы-кривые будут заметно отличаться.

Берестнева Ольга Григорьевна, профессор кафедры математики Института кибернетики ТПУ.

E-mail: ogb6@yandex.ru Область научных интересов: математическое моделирование, информационные технологии.

Воловоденко Виталий Алексеевич, канд. техн. наук, доцент кафедры оптимизации систем управления Института кибернетики ТПУ.

E-mail: volcowvav@tpu.ru Область научных интересов: информационные технологии.

Шаропин Константин Александрович, канд. техн. наук, доцент кафедры прикладной математики Института кибернетики ТПУ.

E-mail: kashar@mail.ru Область научных интересов: компьютерные технологии, математическое моделирование.

Гергет Ольга Михайловна, канд. техн. наук, доцент кафедры прикладной математики Института кибернетики ТПУ.

E-mail: olgagerget@mail.ru Область научных интересов: компьютерные технологии и математическое моделирование.

В рассматриваемом случае наиболее общей формой представления данных является вектор конечномерного пространства R_n

$$A = (a_0, a_1, a_2, \dots, a_{n-1}) \in R_n \quad (1)$$

Для перехода от данного вектора к визуальному образу будет использоваться базис ортонормированных функций $\{\varphi_i(\tau)\}_{i=0}^{\infty}$. В качестве такого базиса можно использовать известные функции, в частности ортонормированные полиномы Лежандра на отрезке $[0,1]$, множество которых мы обозначим через $\{l_i(\tau)\}_{i=0}^{\infty}$.

В таком случае точке с координатами $A = (a_0, a_1, a_2, \dots, a_{n-1})$ можно поставить в соответствие функцию

$$F_A(\tau) = \sum_{i=0}^{n-1} a_i l_i(\tau) \quad (2)$$

Формирование вектора A связано с преобразованием данных. Для характеристики наблюдаемого многомерного объекта большую роль играют значения его координат. В большинстве случаев каждый показатель имеет свою единицу измерения, и его значение будет влиять на вид функции $F_A(\tau)$. Для того чтобы исключить влияние разноименности показателей на вид функции $F_A(\tau)$, необходимо перейти к безразмерным единицам одним из известных способов.

Следует отметить, что порядок включения показателей в вектор A также будет влиять на вид функции $F_A(\tau)$. Для обоснования порядка следования показателей при решении конкретных прикладных задач предлагается проведение экспертного оценивания информативности показателей.

Разница между формулами (1) и (2) заключается в том, что для вектора A из (1) возможно только аналитическое представление, в то время как для функции $F_A(\tau)$ возможно представление в виде графика этой функции. Между (1) и (2) устанавливается однозначная связь в обе стороны, т. е. взаимно-однозначная связь. Если ввести в рассмотрение второй вектор

$$B = (b_0, b_1, b_2, \dots, b_{n-1}) \in R_n,$$

то ему ставится в соответствие функция

$$F_B(\tau) = \sum_{i=0}^{n-1} b_i l_i(\tau).$$

Будем считать, что функции $F_A(\tau)$ и $F_B(\tau)$ являются визуальными образами точек A и B , принадлежащих пространству R_n (рис. 1).

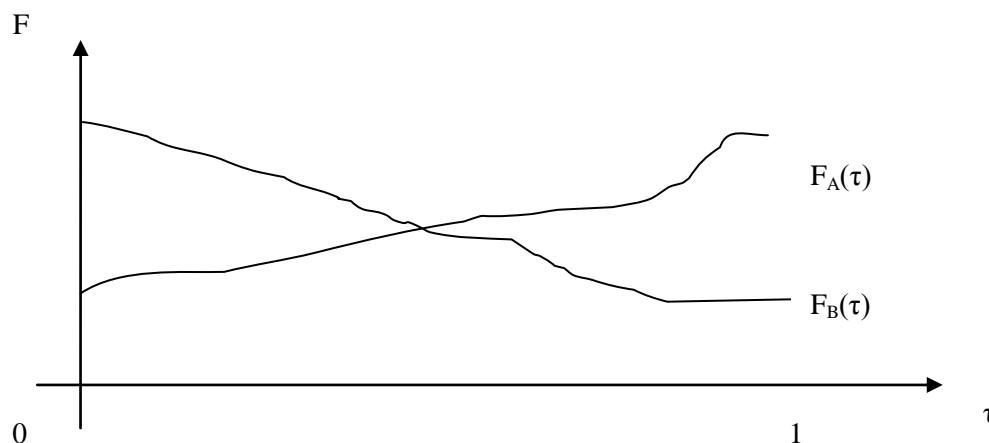


Рис. 1. Визуальные образы точек A и B , принадлежащих пространству R_n

Если с векторами A и B связать точки их концов, а начала векторов разместить в начале координат пространства R_n , то получается взаимно-однозначная связь между точками пространства R_n и функциями-образами $F_A(\tau)$ и $F_B(\tau)$. На рис. 2 введены следующие обозначения для осей координат: $X_0, \dots, X_{n-2}, X_{n-1}$, которые соответствуют показателям $a_0, a_1, a_2, \dots, a_{n-1}$.

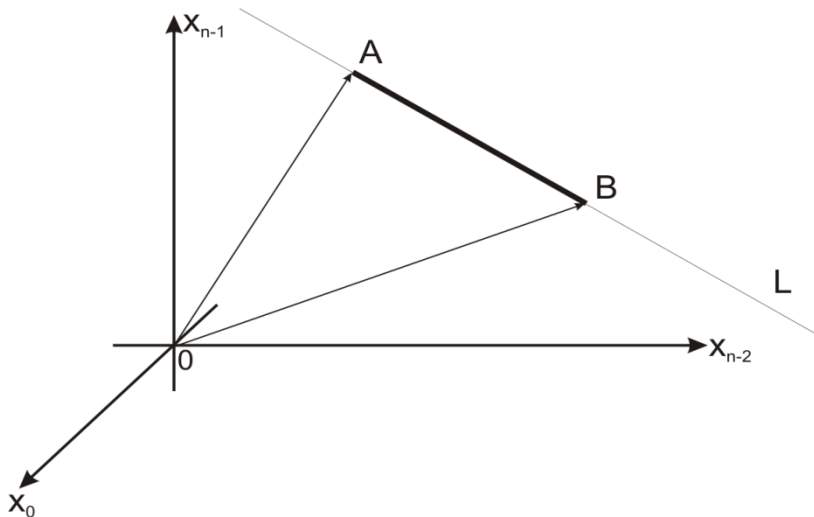


Рис. 2. Условное представление прямой L , проходящей через точки A и B в пространстве R_n

Введем переменную z и образуем линейное выражение

$$C(z) = (1-z)A + zB = ((1-z)a_0 + zb_0, (1-z)a_1 + zb_1, \dots, (1-z)a_{n-1} + zb_{n-1}) \quad (3)$$

Очевидно:

$$C(0)=A \text{ и } C(1)=B.$$

Эта возможность дает право на рассмотрение функции $C(z)$ как функции, представляющей многомерную прямую линию, проходящую через точки A и B в пространстве R_n .

Следовательно, мы можем рассматривать выражение аналогичное (3) для представления отрезка AB :

$$AB = (1-z)A + zB \quad (4)$$

$$z \in [0,1]$$

Параметр z является последовательным параметром и может представлять различные расстояния. Из выражения (4) следует, что можно образовать функцию

$$F_{AB}(z) = \sum_{i=0}^{n-1} (1-z)a_i l_i(\tau) + zb_i l_i(\tau) \quad (5)$$

Обратим внимание, что эта функция будет зависеть от двух аргументов $\{z, \tau\}$. Это очень важно, так как дает возможность получения графика функции $F_{AB}(z) = F_{AB}(z, \tau)$, который является визуальным образом отрезка AB (рис. 3). Таким образом, получается, что на квадрате $[0,1] \times [0,1]$ можно построить гладкую поверхность по формуле (5), которая будет однозначно соответствовать аналитическому выражению (4), представляющему многомерный отрезок AB .

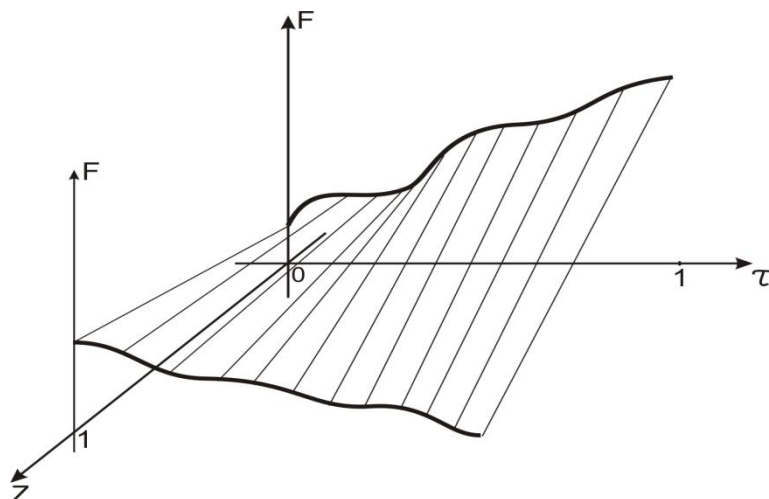


Рис. 3. Визуальное представление поверхности, соответствующей отрезку AB

Рассмотрим пример. Пусть даны многомерные объекты со следующими характеристиками:

$$H1 = \{1, 0, 0, 0\}, H2 = \{0, 1, 0, 0\}, H3 = \{0, 0, 1, 0\}, H4 = \{0, 0, 0, 1\}.$$

Преобразуем их с помощью полиномов:

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow 1 \cdot l_0(\tau) + 0 \cdot l_1(\tau) + 0 \cdot l_2(\tau) + 0 \cdot l_3(\tau)$$

$$\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \Rightarrow 0 \cdot l_0(\tau) + 1 \cdot l_1(\tau) + 0 \cdot l_2(\tau) + 0 \cdot l_3(\tau)$$

$$\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \Rightarrow 0 \cdot l_0(\tau) + 0 \cdot l_1(\tau) + 1 \cdot l_2(\tau) + 0 \cdot l_3(\tau)$$

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \Rightarrow 0 \cdot l_0(\tau) + 0 \cdot l_1(\tau) + 0 \cdot l_2(\tau) + 1 \cdot l_3(\tau)$$

В общем случае:

$$F = \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \end{pmatrix} \Rightarrow f_0 \cdot l_0(\tau) + f_1 \cdot l_1(\tau) + f_2 \cdot l_2(\tau) + f_3 \cdot l_3(\tau) = (F, l(\tau)) = F_1(\tau)$$

Параметр τ представляет собой параметр композиции образа, он безразмерен и играет роль аргумента полиномов. Вектор F нельзя изобразить в трехмерном пространстве, поэтому он заменяется на $F_1(\tau)$, который в свою очередь можно легко представить в виде двумерной кривой.

Для иллюстрации приведем еще один пример из [5]. Предположим, есть два 10-мерных наблюдения A и B со следующими значениями:

A : {53.78, 1, 17.56, 2.54, 6.36, 0.16, 4.63, 8.1, 3.28, 1.9},

B : {50.53, 1.4, 19.05, 2.34, 5.95, 1.53, 3.63, 7.82, 2.98, 2.48}.

Следующие кривые являются визуальными представлениями наблюдений A и B

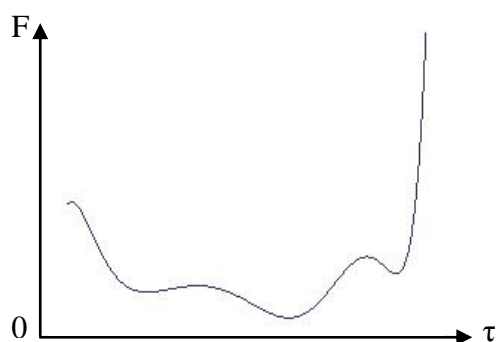


Рис. 4. Кривая A

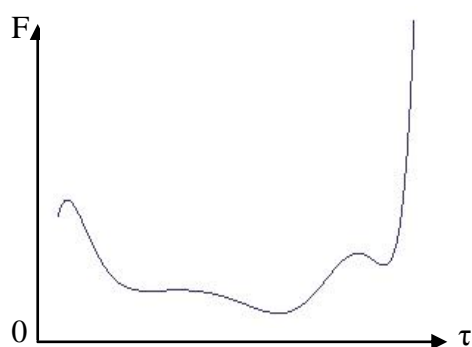


Рис. 5. Кривая B

Теперь совместим эти два образа. Кривые A и B очень похожи друг на друга. Это означает, что исходные наблюдения тоже очень близки друг к другу.

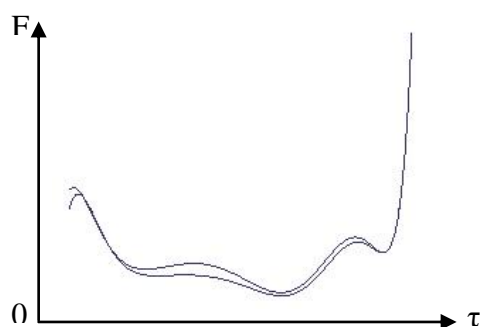


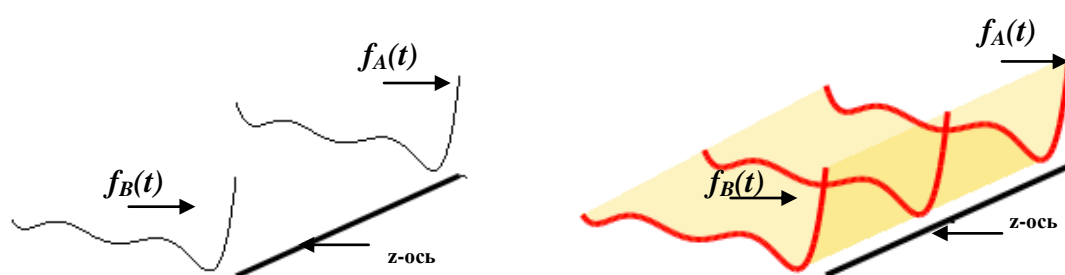
Рис. 6. Кривые A и B

Чем больше кривые неотличимы друг от друга, тем идентичнее наблюдения, которые они представляют, т. е. метод устанавливает взаимно-однозначное соответствие между строками в наборе данных и их кривыми.

Если отобразить кривые наблюдений в трехмерном пространстве, используя третье измерение, называемое также «Z-измерением», как расстояние в многомерном пространстве или промежуток времени между двумя наблюдениями, можно обнаружить много интересных свойств.

Введение понятия расстояния между наблюдениями позволяет отображать наблюдения в трехмерном пространстве, отображая значения величины расстояния на «z-оси». Метрика расстояния выбирается произвольно и соответствует, например, геометрическому (Евклидовое), статистическому (Махаланобис) или хронологическому (временной промежуток) расстояниям. Таким образом появляется возможность отображения как статических, так и динамических данных.

В силу линейности преобразования $A \leftrightarrow f_A(t)$, образу отрезка прямой между многомерными наблюдениями A и B соответствует поверхность, соединяющая образы наблюдений. Любая кривая, представляющая наблюдение с промежуточными значениями, будет лежать на этой поверхности, как показано на рис. 7.



Образы наблюдений A и B в трехмерном изображении Образ отрезка между наблюдениями A и B с промежуточным наблюдением

Рис. 7. Изображение отрезка многомерной прямой [5]

Для более детального сравнения наблюдений, особенно в случае неоднородных единиц измерения показателей, данные можно трансформировать одним из традиционных способов:

- *нормализация* – для выражения результатов в единой системе измерений
- *стандартизация* – для сравнения переменных и/или наборов данных с различными характеристиками распределения или единицами измерения

Представленный подход реализован в пакете *NovoSpark Visualizer*, на базе которого авторами был успешно решен ряд прикладных задач анализа и интерпретации многомерных данных в медицине [6, 7], педагогике [4] и социальной сфере [6, 9].

Таким образом, представление *многомерного* наблюдения в виде *двумерного* образа (кривой) гарантирует, что близким по значениям наблюдениям A и B будут соответствовать визуально близкие образы-кривые; для сильно различающихся по значениям наблюдений их образы-кривые будут заметно отличаться. Становится возможным автоматически классифицировать наблюдения, определять наиболее важные переменные в модели, производить кластеризацию данных, визуально сравнивать индивидуальные наблюдения и целые наборы данных, а также выполнять много других задач в работе с многомерными данными.

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

СПИСОК ЛИТЕРАТУРЫ

1. Шаропин К.А., Берестнева О.Г., Шкатова Г.И. Визуализация результатов экспериментальных исследований //Известия Томского политехнического университета, 2010. – Т. 316. – № 5. – С. 172–176.
2. Дюк В.А., Эммануэль В. Информационные технологии в медико-биологических исследованиях. – СПб: Питер, 2003. – 528 с.

3. Попечителей Е.П. Анализ числовых таблиц в биотехнических системах обработки экспериментальных данных. – Л.: Наука, 1985. – 148 с.
4. Марухина О.В., Берестнева О.Г., Воловоденко В.А., Шаропин К.А. Технологии визуализации результатов экспериментальных исследований // Информационные и математические технологии в науке и управлении: Труды XVI Байкальской Всерос. конф.. Часть 3. – Иркутск, 2010. – С. 165–171.
5. 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.
6. Берестнева О.Г., Пеккер Я.С., Шаропин К.А., Воловоденко В.А. Выявление скрытых закономерностей в медицинских и социально-психологических исследованиях // Аппликативные вычислительные системы: Труды 2-й Междунар. конф. по аппликативным вычислительным системам. – Москва, 29–31 октября 2010. – Москва: Институт Актуального образования «ЮрИнфоР-МГУ», 2010. – С. 287–296.
7. Марухина О.В., Берестнева О.Г., Шаропин К.А., Осадчая И.А. Когнитивная графика в социально-психологических исследованиях // Информационные и математические технологии в науке и управлении: Труды XVI Байкальской Всерос. конф. «Информационные и математические технологии в науке и управлении». Часть 3. – Иркутск, 2011. – С. 176–181.

Поступила 15.10.2011 г.