
НОВОЕ В ЭМПИРИЗМЕ

В.М. МОРДАШЕВ

НИЦ «Курчатовский институт» (г. Москва)
mrdshv@rambler.ru

Препятствием к осмыслению большинства численных, прежде всего экспериментальных данных, встречающихся в практике исследований, является их многомерность. Мы живем в трехмерном мире и не способны визуально воспринимать данные, зависящие более чем от двух переменных. В статье обсуждается созданный автором нелинейный многофакторный анализ, позволяющий приближенно визуализировать данные, зависящие от произвольного числа переменных. Метод разработан как для упорядоченных данных типа многомерных матриц (активного эксперимента), так и для случайного набора данных (пассивного эксперимента).

Ключевые слова: приближение функций многих переменных, многофакторный анализ, визуализация многомерных данных, эмпиризм.

1. ЭКСПЕРИМЕНТ И ТЕОРИЯ. ПРОБЛЕМА МНОГОМЕРНЫХ ДАННЫХ

Что важнее – теория или эксперимент?

Вопрос кажется глупым: знания, науки без теории и экспериментов не бывают! И тем не менее эволюция взглядов на них такова:

Исаак Ньютон (английский физик, 1642–1727): «Гипотез не измышляю, ибо все то, что не может быть выведено из явлений, должно быть названо гипотезой».

Михаил Ломоносов (русский физик, 1711–1765): «Из наблюдений устанавливать теорию, через теорию направлять наблюдения – есть лучший способ к отысканию Истины».

Петр Капица (русский физик, 1894–1984): «Отрыв теории от эксперимента, опыта, практики наносит ущерб прежде всего теории».

Василий Ключевский (русский историк, 1841–1911): «Факт, не приведенный в концептуальную схему, есть лишь смутное представление, из которого нельзя сделать научного употребления».

Фернанд Бродель (французский историк, 1902–1985): «Факты – это лишь пыль и являются в истории лишь краткими вспышками».

В результате подобной эволюции эксперимент сегодня уже не первичный источник знаний, а, главным образом, средство подтверждения или опровержения гипотез и теорий.

Современный стандартный сценарий получения новых знаний таков: сначала гипотеза или теория, затем подтверждение (или опровержение) их экспериментом.

Ушел в прошлое сценарий, который обозначил М. Ломоносов и по которому создавалась, например, молекулярно-кинетическая теория вещества: от эмпирических законов Бойля–Мариотта (1662 г., 1676 г.) – зависимости объема газа от давления при постоянной температуре, Шарля (1787 г.) – зависимости давления газа от температуры при постоянном объеме, Гей-Люссака (1802 г.) – зависимости объема газа от температуры при постоянном давлении к эмпирическому закону Клапейрона (1834 г.), объединяющему эти три эмпирических закона, затем к теоретическому закону Менделеева–Клапейрона (1874) и, наконец, к самой молекулярно-кинетической теории вещества.

Современное соотношение теории и эксперимента беспокоило многих выдающихся ученых, в том числе Тесла и Капицу.

«Сегодняшние ученые заменили математикой эксперименты, и они блуждают прочь через уравнение после уравнения, и, в конечном счете, строят структуру, которая не имеет никакого отношения к действительности» – Никола Тесла.

«Два зачина – от опыта или теории на глубоком уровне отражают два подхода к миру. Для нашей страны примат теории над опытом, практикой, особенно в социальной, так и в технической и научной сферах стал поистине роковым и источником наших многих бед. Преодоление стереотипа есть, быть может, важнейшая задача перестройки нашего мышления» – П.Л. Капица.

Несомненно, главной причиной ухода от эмпирического подхода к теоретическому, умозрительному, является многомерность изучаемых явлений и эмпирических данных, их зависимость от многих факторов.

Мы живем и наблюдаем, «видим» в трехмерном мире, а в четырехмерном и более видеть не способны!

Именно поэтому не известны эмпирические законы, содержащие более двух независимых переменных факторов. А те, что содержат и называются эмпирическими, на самом деле теоретические модели с подобранными эмпирическими коэффициентами, благо, что сегодня вычислительная математика и современные компьютеры позволяют находить наилучшие с точки зрения аппроксимации эмпирические коэффициенты для сколь угодно сложной теоретической модели. Теоретические модели с такими коэффициентами правильнее называть «полуэмпирическими». Такое название ранее и существовало, но за «ненужностью» вышло из употребления.

Возможность решения проблемы чувственного ощущения и осмысления многомерных данных «в лоб» сомнительна. Цвет, звук, запах, наверное, могут как-то расширить наши возможности, но стать столь же эффективными, как графика, они вряд ли смогут.

Альтернативным (а может быть, и единственным) подходом к решению является представление многомерной, многофакторной зависимости связной комбинацией, системой зависимостей меньших размерностей, т.е. представление ее некоторым функционалом с переменными, разделяющимися на группы по одной-две переменные.

Если такой функционал известен, то он может быть визуализирован, т.е. представлен графически. Этим процессом занимается номография.

Первая номограмма была построена Пуше (Pouchet) в 1795 г. для решения уравнения $x \cdot y = z$. Первую систему построения номограмм дал Лалан (Lalanne) в 1843 г. Основания общей теории номографических построений даны французским ученым Оканем (*D'Ocagne*) в 1884–1891 гг. Первым в России вопросами номографии начал заниматься Н.М. Герсегонов в 1906–1908 гг. В прошлом номограммы (самая распространенная из них – логарифмическая линейка) играли огромную роль в различных областях науки и особенно техники. Трудно указать область техники, где они бы не использовались.

Основное назначение номограммы – вычислительный прибор. Поэтому с появлением компьютеров они сошли на нет. Но вычисления – не единственное достоинство номограмм. Номограмма позволяет не только выполнять почти мгновенные вычисления, в том числе и для обратных задач, но и осмысливать и прогнозировать тенденции и зависимости благодаря визуализации.

На рис. 1 приведена номограмма (вернее, ее скелет), связывающая 9 переменных, из которых 8 – независимые.

Но номограмма может быть построена только тогда, когда известны реализуемый ею функционал и разделение переменных в нем. А поскольку

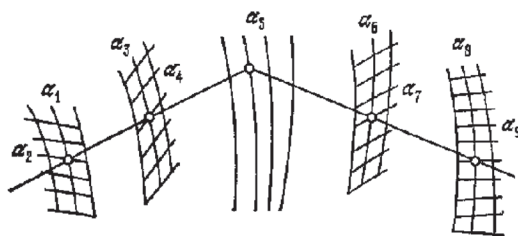


Рис. 1. Номограмма, отражающая взаимозависимость переменных

ку такое для большинства эмпирических (и не только эмпирических) данных неизвестно, то номограммы в практике исследований сошли «со сцены». Так что виной исчезновения номограмм являются не только компьютеры, но прежде всего отсутствие приближенных методов разделения переменных и соответственно выявления номографируемых функционалов для конкретных массивов численных данных.

Между тем определенный оптимизм в том, что разделение переменных возможно, если и не всегда, то в большинстве случаев, порождают результаты академика А.Н. Колмогорова и его учеников В.И. Арнольда [1] и А.Г. Витушкина (впоследствии также академиков), полученные в процессе решения 13-й проблемы Гильберта и касающиеся представления функций многих переменных суперпозициями функций меньшего числа переменных.

Теорема А.Н. Колмогорова [2] гласит: любая непрерывная в многомерном единичном кубе функция многих переменных может быть представлена суммой функций, каждая из которых зависит от суммы функций, зависящих от одной переменной:

$$\varphi(x_1, x_2, \dots, x_n) = \sum_{i=1}^{2n+1} f_i \left\{ \sum_{j=1}^n \xi_{ij}(x_j) \right\} \quad (1)$$

Поскольку это утверждение верно для непрерывных функций, то оно тем более верно и для дискретных функций (например, многомерных матриц), поскольку каждая из них является подмножеством бесконечного множества непрерывных функций со значениями, совпадающими со значениями дискретной функции.

2. НЕЛИНЕЙНЫЙ МНОГОФАКТОРНЫЙ АНАЛИЗ

Ниже речь пойдет о нелинейном многофакторном анализе (НЛМФА) [3].

В его основе лежит теорема автора [4], гласящая: среди линейных комбинаций линейно независимых в многомерном параллелепипеде функций (на практике многомерных матриц) $\varphi(x_1, x_2, \dots, x_n)$, приближаемых суммой функций меньшего числа переменных $\sum_{l=1}^m f_l(\tau^l)$ (матриц

меньшей размерности), где $\tau^l \in \{x_1, x_2, \dots, x_n\}$, существует, как правило, единственная такая, что при нормировке коэффициентов a_i , $aNa^T=1$, где N – положительно определенная матрица, реализуется наименьшее среднеквадратичное отклонение при приближении

$$\sum_{i=1}^k \alpha_i \varphi_i(x_1, x_2, \dots, x_n) \approx \sum_{l=1}^m f_l(\tau^l) \quad (2)$$

Другими словами, как только задана система многомерных матриц $\varphi_i(x_1, x_2, \dots, x_n)$ и комбинация множеств переменных τ^l , коэффициенты a_i и правая часть (2), $\sum_{l=1}^m f_l(\tau^l)$, определяются однозначно.

2.1. СЛУЧАЙ ОДНОЙ МНОГОМЕРНОЙ МАТРИЦЫ, «ПОЛНЫЙ ФАКТОРНЫЙ ЭКСПЕРИМЕНТ»

Если в качестве системы линейно независимых функций $\varphi_i(x_1, x_2, \dots, x_n)$ рассматривать систему линейно независимых известных преобразований одной функции $\varphi(x_1, x_2, \dots, x_n)$, например ее степеней, то НЛМФА находит такое нелинейное преобразование L (новую шкалу измерения) анализируемой функции, что оно (преобразование) наилучшим образом (в смысле метода наименьших квадратов) аппроксимируется суммой функций меньшего числа переменных:

$$L(x_1, x_2, \dots, x_n) = L\{\varphi(x_1, x_2, \dots, x_n)\} = \sum_j^k a_j \varphi^j(x_1, x_2, \dots, x_n) \approx \sum_{l=1}^m f_l(\tau^l), \quad (3)$$
 где τ^l – некоторое подмножество множества переменных $\{x_1, x_2, \dots, x_n\}$, а коэффициенты a_i нормированы так, чтобы обеспечить примерное равенство стандартных погрешностей аппроксимаций (2) и (3).

Очевидно, что имея такое приближение, нетрудно при необходимости получить и приближение самой приближаемой функции, например, с помощью стандартных итерационных методов:

$$\varphi(x_1, x_2, \dots, x_n) = L^{-1}\left\{\sum_{l=1}^m f_l(\tau^l)\right\}, \quad (4)$$

где L^{-1} – функция, обратная функции L .

Как только указана степень полинома, требования к монотонности новой шкалы измерения (для взаимной однозначности $L(\varphi)$ и $\varphi(L)$) и комбинация множеств переменных $\tau^1, \tau^2, \dots, \tau^l$, коэффициенты полинома и приближающая сумма функций меньшего числа переменных определяются однозначно.

Рассматривая различные комбинации множеств переменных $\tau^1, \tau^2, \dots, \tau^l$ в правой части уравнения (3) и степени полинома k в левой, можно выбрать такую аппроксимацию, которая представляется наиболее простой, но удовлетворительной по точности приближения.

Правая часть визуализируема и номографируема, если множества τ^l содержат не более двух переменных.

ПРИМЕР 1

Данный пример является типичным случаем применения НЛМФА для многомерной таблицы данных с дискретными факторами и визуализации ее в виде приближенной номограммы.

В статье [5], посвященной физиологии зрения, приведены экспериментальные данные (табл. 1) числа порогов цветоразличения, φ , у лиц с различными типами зрения, x_4 , в различных зонах спектра, x_1 , после предварительной адаптации при различном цвете, x_2 , с различной его насыщенностью, x_3 . M – среднее число порогов цветоразличения, m – среднеквадратичное отклонение среднего.

В табл. 2 приведены результаты применения НЛМФА к табл. 1 и процесс выбора адекватной модели по среднеквадратичному отклонению.

Таблица 1

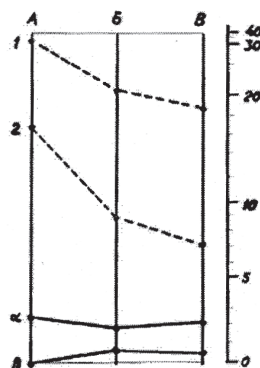
Число порогов цветоразличения, φ

Условия наблюдений			Тип зрения, x_4		
Зона спектра, нм, x_1	Цвет при предварительной адаптации, λ , нм, x_2	Насыщение цвета адаптации, %, x_3	Нормальная трихромазия, $M \pm m$	Дейтераномалия типа В, $M \pm m$	Протаномалия типа В, $M \pm m$
665–620	650	100	20±0,72	16±0,51	14±0,43
		30	30±1,23	19±0,31	18±0,1
	540	100	23±0,73	17±0,31	14±0,13
		30	32±1,1	19±0,12	18±0,61
540–520	650	100	12±0,6	6±0,4	5±0,2
		30	17±1,1	7±0,41	6±0,2
	540	100	12±0,16	6±0,14	5±0,43
		30	17±0,53	8±0,33	6±0,4

Таблица 2

Результаты аппроксимации НЛМФА

Разделение переменных на группы в приближающей сумме $\sum_{i=1}^m f_i(\tau^i)$	Шкала $L(\varphi)$ (полином)	σ	ч.с.с.	Принято?
$\{x_1, x_2\}, \{x_1, x_3\}, \{x_1, x_4\}, \{x_2, x_3\}, \{x_2, x_4\}, \{x_3, x_4\}$	1,406 $\varphi - 0,01488 \varphi^2$	0,617	8	Да
$\{x_1, x_2\}, \{x_1, x_3\}, \{x_1, x_4\}, \{x_2, x_3\}, \{x_2, x_4\}$	1,438 $\varphi - 0,01614 \varphi^2$	0,615	9	Да
$\{x_1, x_2\}, \{x_1, x_3\}, \{x_1, x_4\}, \{x_2, x_3\}$	1,577 $\varphi - 0,02182 \varphi^2$	0,659	10	Да
$\{x_1, x_2\}, \{x_1, x_3\}, \{x_1, x_4\}$	1,286 $\varphi - 0,01028 \varphi^2$	0,952	12	Нет
$\{x_1, x_4\}, \{x_2, x_4\}, \{x_3, x_4\}$	1,577 $\varphi - 0,02182 \varphi^2$	0,632	11	Да
$\{x_1, x_2\}, \{x_1, x_4\}, \{x_3\}$	1,577 $\varphi - 0,02182 \varphi^2$	0,608	13	Да
$\{x_1, x_4\}, \{x_2\}, \{x_3\}$	1,577 $\varphi - 0,02182 \varphi^2$	0,881	15	Нет
$\{x_1, x_2\}, \{x_1, x_4\}$	1,577 $\varphi - 0,02182 \varphi^2$	0,635	14	Да
$\{x_1, x_4\}, \{x_3, x_4\}$	φ	1,23	15	Нет



Число порогов цветоразличения равно расстоянию по шкале между точками «насыщение цвета при адаптации» и «зона спектра» для выбранного типа зрения.

Тип зрения: А – нормальная трихромазия;

Б – дейтераномалия типа В;

В – протаномалия типа В.

Зона спектра: 1 – 620–665 нм;

2 – 520–540 нм.

Насыщение цвета при адаптации: α – 100 %;

β – 30 %.

Рис. 2. Номограмма для числа порогов цветоразличения

При этом учитывалось, что если приближение принято для некоторой комбинации переменных и степени полинома, то оно тем более было бы принято для более сложной комбинации и большей степени полинома. В качестве начальной принята $0,1 < \partial L / \partial \varphi < 10$ степень полинома 2, допустимая монотонность шкалы принята для всего диапазона исходных значений φ .

В табл. 2: σ – несмещенное среднееквадратичное отклонение модели от данных, ч.с.с. – число степеней свободы – разница между числом данных и числом определяемых параметров модели (числом связей модели).

Функции $f_1(x_1, x_4)$ и $f_2(x_3, x_4)$ приведены в виде табл. 3–4.

Таким образом, найдено адекватное (со стандартной ошибкой 0,5, ~5 %) представление исходной таблицы данных в виде выражения:

$$L(\varphi) = 1,577\varphi - 0,02182\varphi^2 \approx f_1(x_1, x_4) + f_2(x_3, x_4), \quad (5)$$

которое представляется номограммой, приведенной на рис. 2.

Таблица 3

Значения функции $f_1(x_1, x_4)$			
Зона спектра, нм, x_1	Вид зрения, x_4		
	Нормальная трихромазия	Дейтераномалия типа В	Протаномалия типа В
665–620	23,60	20,25	18,52
540–520	16,21	8,82	6,94

Таблица 4

Значения функции $f_2(x_3, x_4)$			
Насыщение цвета адаптации, %, x_3	Вид зрения, x_4		
	Нормальная трихромазия	Дейтераномалия типа В	Протаномалия типа В
100	0,00	0,00	0,00
30	4,22	1,96	2,41

Забавно отметить, что авторы статьи [5] при традиционном анализе данных не заметили независимости от цвета при предварительной адаптации, x_2 . Очевидно, это противоречило их теоретическим воззрениям.

2.1.1. НАКОПЛЕННЫЙ ОПЫТ

Ниже приведены еще два примера применения НЛМФА к данным полного факторного эксперимента.

ПРИМЕР 2

В книге [6] приведены статистические данные о смертности в различных государствах бывшего СССР для различных групп населения в различные годы, трехмерная матрица: *Death* (Государство, Группа населения, Время).

Результат применения к ней НЛМФА дает модель $Death = D \cdot Y,$ (6) где *D* и *Y* приведены на рис. 3–4.

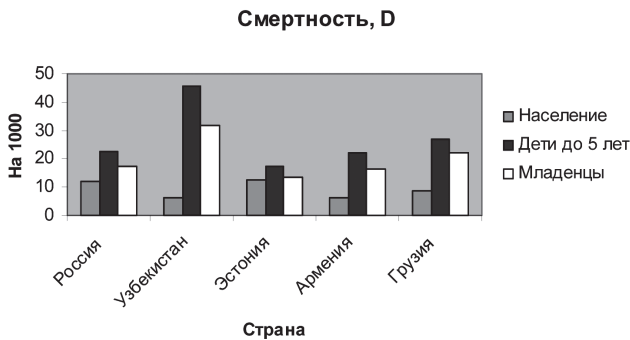


Рис. 3. Смертность по разным странам и группам населения в 1990 г.

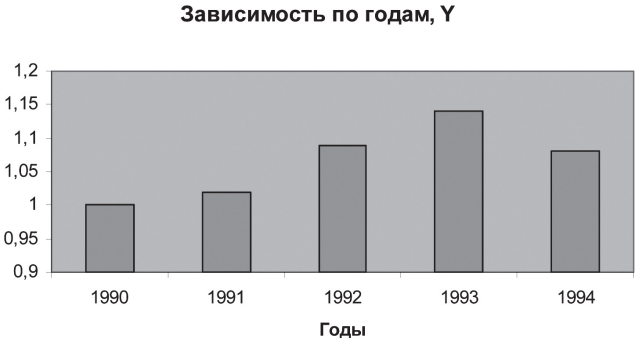


Рис. 4. Коэффициент Y в различные годы

Модель можно считать удовлетворительной, так как при размахе исходных данных в 8,5 раза стандартная ошибка модели составляет 8 %.

ПРИМЕР 3

НЛМФА был применен к анализу экспериментальных данных по моделированию аварийного охлаждения активной зоны ядерного реактора. Эти данные не совпадали с существовавшими теоретическими моделями не только количественно, но и качественно.

Для важнейшей характеристики изучаемого процесса – скорости фронта пар–вода – НЛМФА позволил получить модель в виде произведения функций с переменными, разделяющимися на группы, содержащие по одной переменной [7]. Для точечных зависимостей от этих переменных, без каких-либо физических соображений, а только из соображений простоты и точности были подобраны описывающие их функции. В результате была построена аналитическая эмпирическая формула:

$$u = 0,37 \cdot w^{0,7} \cdot (1 - T/1200) \cdot (1 - 0,07 \cdot \sqrt{\Delta T}) \text{ [см/с]}, \quad (7)$$

с достаточно высокой точностью (стандартная ошибка 6 % при размахе данных в 7,5 раза) описывающая скорость фронта пар–вода, u , в широком диапазоне значений влияющих факторов:

массовая скорость, w 10–75 кг/м²с,
 начальная температура стенки, T 600–1100 К,
 недогрев воды на входе, ΔT 5–80 К.

За время разработки и существования НЛМФА было рассмотрено около 200 задач трех-четырёхмерных матриц из различных областей науки и техники: физики и техники атомных реакторов, теплофизики и теплотехники, радиохимии, материаловедения, физиологии, экономики, демографии и т.п. Для всех, кроме одного случая, были найдены удовлетворительные по точности визуализированные и/или аналитические модели.

Исключение составила трехмерная матрица статистического критерия Фишера – сравнения двух выборок по среднеквадратичным отклонениям при различных уровнях значимости и степенях свободы.

Накопленный опыт позволяет утверждать, что НЛМФА эффективен, по крайней мере, в 99 % случаев.

2.2. ОРТОГОНАЛЬНЫЕ ПЛАНЫ

Как нетрудно было заметить, НЛМФА для многомерных матриц требует достаточно большого объема спланированных данных для всех сочетаний значений факторов, что позволяет находить приближения (3) вплоть до случая, когда комбинация переменных в правой части состоит из множеств, содержащих по $n-1$ переменные. Такой объем представля-

ется излишне большим, если речь идет о планировании экспериментов с целью получения приближений вида (3).

НЛМФА разработан и для так называемых ортогональных таблиц, таких как «латинские квадраты». В них каждое сочетание значений любых двух факторов встречается обязательно и только один раз. В ортогональных кубах встречается обязательно и только один раз каждое сочетание любых трех факторов и т.д.

Рассмотрим пример с латинским квадратом.

ПРИМЕР 4

Пусть некоторая функция $F(x, y, z)$ зависит от трех независимых переменных x, y, z как $F = \sqrt{x^2 + y^2 + z^2}$, но мы об этом не знаем. Наша задача – как можно точнее выявить эту зависимость в области $D = \{x, y, z = \{0, 2, 4\}\}$ заданной в виде латинского квадрата (табл. 5).

Результаты анализа табл. 5 с помощью НЛМФА приведены в табл. 6 (σ – несмещенное среднеквадратичное отклонение модели от данных; ч.с.с. – число степеней свободы – разница между числом данных и числом определяемых параметров модели).

Результирующее приближение имеет вид

$$L(x, y, z) = 0,0976 \cdot F + 0,1017 \cdot F^2 \approx f_1(x) + f_2(y) + f_3(z) = L(x) + L(y) + L(z) - 2L(0),$$

где $L(x), L(y), L(z), L(0)$ представлены на рис. 5.

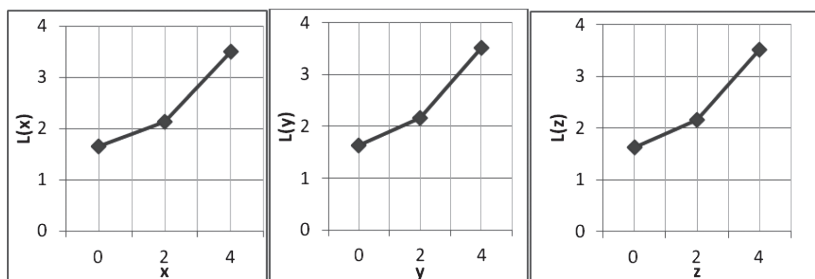
Таблица 5

F(x, y, z)							
x	y	z	F	x	y	z	F
0	0	0	0	2	4	0	4,472
0	2	2	2,828	4	0	4	5,657
0	4	4	5,657	4	2	0	4,472
2	0	2	2,828	4	4	2	6
2	2	4	4,899				

Таблица 6

Результаты анализа табл. 5 с помощью НЛМФА

Разделение переменных на группы в приближающей сумме	$L\{F(x, y, z)\}$	σ – ср.кв. погрешн.	ч.с.с. – число степеней свободы	Приближение принято?
{x}, {y}, {z}	$0,0976 \cdot F + 0,1017 \cdot F^2$	0,095	1	Да
{x}, {y}	$0,0976 \cdot F + 0,1017 \cdot F^2$	1,37	3	Нет
{x}, {z}	$0,0976 \cdot F + 0,1017 \cdot F^2$	1,37	3	Нет
{y}, {z}	$0,0977 \cdot F + 0,1017 \cdot F^2$	1,37	3	Нет
{x}, {y}, {z}	F	0,686	2	Нет

Рис. 5. Значения функций $L(x)$, $L(y)$, $L(z)$, $L(0) = 2,434$

Поскольку новая шкала, найденная НЛМФА, имеет вид

$$b_1 \cdot F(x, y, z) + b_2 \cdot F^2(x, y, z) = L(x, y, z), \text{ то} \quad (8)$$

$$F = \frac{-b_1 \pm \sqrt{b_1^2 - 4b_2L}}{2b_2}.$$

Из рис. 5 следует, что $L(x)$, $L(y)$, $L(z)$ являются, скорее всего, полиномами 2-й степени. Поэтому, с учетом (6), функцию F будем искать в виде

$$F = \sqrt{a_1x + a_2x^2 + a_3y + a_4y^2 + a_5z + a_6z^2} + a_7 + a_8. \quad (9)$$

Приближение табл. 5 формулой (9) (программы для такого приближения существуют во многих стандартных пакетах) со стандартной ошибкой $\sigma \approx 0,0001$ дает формулу $F = \sqrt{x^2 + y^2 + z^2}$, т.е. выявляет истинную закономерность.

2.3. ПЛАНЫ С «ДЫРКАМИ»

НЛМФА был разработан также для многомерных матриц с отсутствием данных в их некоторых, вообще говоря, случайных ячейках.

3. НЛМФА И «ПАССИВНЫЙ ЭКСПЕРИМЕНТ» – НЕУПОРЯДОЧЕННЫЙ МАССИВ ДАННЫХ

Все, что говорилось выше, относилось к упорядоченным данным, активному эксперименту, когда в многомерных данных независимые переменные принимают вполне определенные заранее назначенные значения. Но на практике такого рода данные встречаются относительно редко. Преобладают обычно наблюдения над неуправляемыми или недостаточно управляемыми объектами, неупорядоченные массивы данных.

Примерами областей, где преобладают неупорядоченные массивы данных, пассивный эксперимент, могут служить астрономия, экология.

Как быть в этих случаях?

Первое, что приходит в голову, – из массива неупорядоченных данных путем интерполяции и экстраполяции построить правдоподобную многомерную упорядоченную таблицу, с помощью НЛМФА выявить разделение независимых переменных и тип шкалы, с их помощью предположить вид аппроксимирующей функции с неизвестными коэффициентами и найти эти коэффициенты, аппроксимируя этой функцией исходные неупорядоченные данные.

Рассмотрим этот подход на примере.

ПРИМЕР 5

Пусть источником данных является та же функция $F = \sqrt{x^2 + y^2 + z^2}$, что и в примере 2, а переменные x, y, z принимают случайные значения, равномерно распределенные на отрезке $[0, 4]$. Построенные таким образом исходные данные представлены табл. 7 (объемом 50 данных).

На рис. 6–8 приведены значения F в зависимости от каждой из переменных при игнорировании значений других переменных, т.е. проекции $F(x, y, z)$ на оси x, y и z соответственно.

Из рис. 6–8 трудно сделать какие-либо выводы о зависимости $F(x, y, z)$.

Для возможности применения НЛМФА построим правдоподобную таблицу данных полного факторного эксперимента.

Для выявления тренда зависимостей от переменных желательно иметь как можно больше точек по этим переменным. Поэтому в качестве правдоподобной таблицы построим трехмерную матрицу $7 \times 7 \times 7$ – табл. 8. Значения каждой переменной для матрицы выберем следующим образом. Из табл. 7 выберем минимальное и максимальное значения переменной и назовем их минимальным и максимальным значением в правдоподобной табл. 8. Остальные значения равномерно распределим между ними.

Для нахождения правдоподобных значений F табл. 8 применим линейную интерполяцию и экстраполяцию следующим образом. Для точки ячейки табл. 8 с соответствующими координатами x, y, z найдем ближайшие (в данном случае 3) точки из табл. 7 и их значения F . Проведем через них гиперплоскость и на ней по координатам ячейки определим значение F для табл. 8.

Погрешности в восстановлении значений F , возникающие при линейной интерполяции, особых тревог не вызывают. Но при экстраполяции возможны существенные погрешности. В данном случае мы попытаемся избежать их тем, что значения F для табл. 8, выходящие за рамки диапазона значений F табл. 7, будем заменять соответственно на максимальное или минимальное значения этого диапазона.

Таблица 7

Случайные реализации функции $F(x, y, z)$

x	y	z	F	x	y	z	F
2,6764	2,8866	2,6005	4,7179	2,3147	0,2888	3,2437	3,9953
1,5392	0,9804	0,6815	1,9480	1,0097	1,8386	1,7616	2,7392
3,1365	0,4032	0,0158	3,1623	1,5039	2,8439	3,3165	4,6205
0,8210	3,6922	1,8866	4,2268	1,0320	0,511	1,1767	1,6465
2,5785	0,8183	2,8719	3,9455	3,7626	1,2241	2,3253	4,5895
0,4519	0,6675	1,3541	1,5759	3,4496	2,2110	2,1921	4,6469
3,7799	3,7546	1,7726	5,6149	3,0154	1,4367	2,9802	4,4765
0,3782	3,8634	0,7055	3,9455	3,1905	0,6602	1,6480	3,6512
3,7438	1,7307	0,8000	4,2013	1,3121	1,8827	1,2491	2,6128
1,0148	1,4688	0,4981	1,8534	3,3073	3,8369	2,1616	5,5076
3,5466	2,3149	1,0769	4,3700	1,5296	0,5563	1,2785	2,0698
0,2733	1,8543	3,2704	3,7695	3,6226	3,3245	2,2506	5,4075
3,6349	0,7826	2,4462	4,4508	2,0570	0,1075	0,5900	2,1427
1,2149	2,9419	1,3548	3,4593	2,0742	1,8437	1,1762	3,0141
3,4283	3,2259	0,1270	4,7092	1,3404	1,8278	1,6303	2,7920
3,8666	1,0442	3,3020	5,1908	1,8673	1,2905	0,2734	2,2863
3,1817	3,0439	1,3687	4,6110	0,3130	3,1428	3,1758	4,4790
3,8757	2,0697	2,0750	4,8590	1,2579	2,4838	2,6377	3,8353
0,3100	0,4594	1,0888	1,2217	2,2449	2,7834	1,2914	3,8020
2,1599	0,1863	0,0922	2,1699	3,3473	1,1541	3,5102	4,9858
3,5094	3,0659	0,2031	4,6644	3,4601	1,2091	1,5704	3,9876
3,4195	2,7636	0,5365	4,4292	2,6376	2,0393	3,5145	4,8444
2,0991	3,1560	2,0025	4,2868	1,9334	2,9513	0,1467	3,5313
2,1045	3,1147	1,8196	4,1763	1,0321	2,8059	1,5724	3,3781
0,3662	1,5288	0,9893	1,8575	2,5861	0,5313	1,9783	3,2991

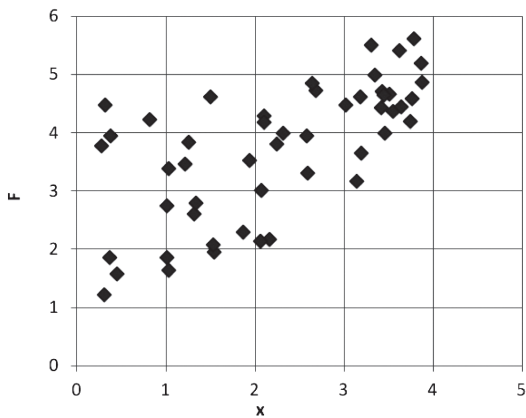


Рис. 6. Проекция $F(x, y, z)$ из табл. 7 на ось x

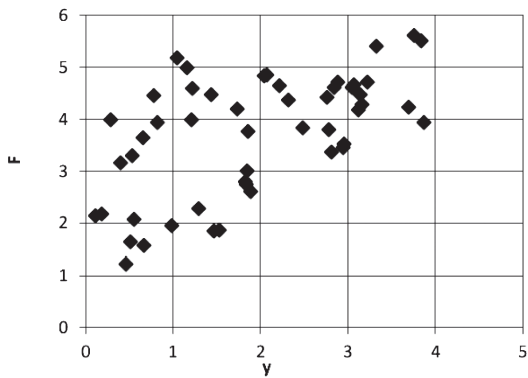


Рис. 7. Проекция $F(x, y, z)$ из табл. 7 на ось y

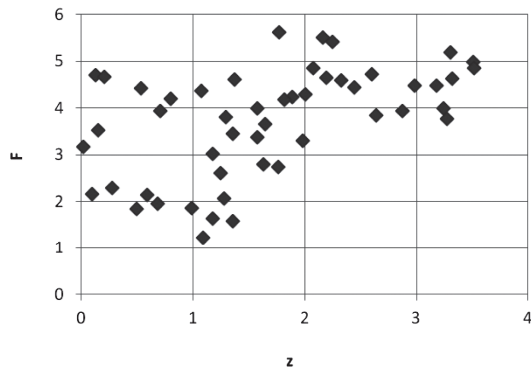


Рис. 8. Проекция $F(x, y, z)$ из табл. 7 на ось z

Поскольку, полученная таким образом табл. 8 имеет достаточно большой объем, здесь она представлена только своими 7-ю первыми и 7-ю последними строками.

Для анализа примем начальную степень полинома 2, а число переменных в разделенных множествах по одной, как в примере 4, так как уже при 2 переменных в разделенных множествах и 7 значениях каждой переменной система была бы неопределенной – число степеней свободы стало бы меньше нуля.

Результат анализа табл. 8 с помощью НЛМФА представлен табл. 9. В ней: σ – несмещенное среднее квадратичное отклонение модели от данных табл. 9; ч.с.с. – число степеней свободы – разница между числом данных табл. 7 и числом определяемых параметров модели (числом связей модели).

Результирующее приближение имеет вид:

$$L(x, y, z) = 0,281 \cdot F + 0,9109 \cdot F^2 \approx f_1(x) + f_2(y) + f_3(z) = L(x) + L(y) + L(z) - 2L(0), \text{ где } L(0) = 2,528, \text{ а } L(x), L(y), L(z) \text{ представлены рис. 9–11.}$$

Поскольку новая шкала, найденная НЛМФА, как и в примере 4, имеет вид $b_1 \cdot F(x, y, z) + b_2 \cdot F^2(x, y, z) = L(x, y, z)$,

Таблица 8

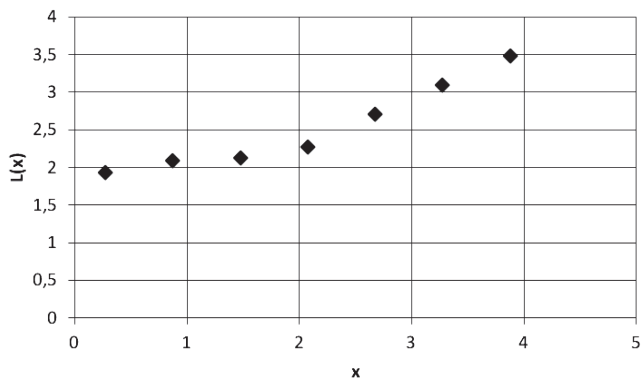
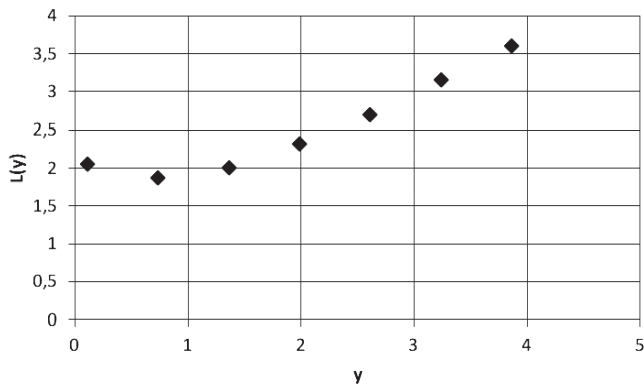
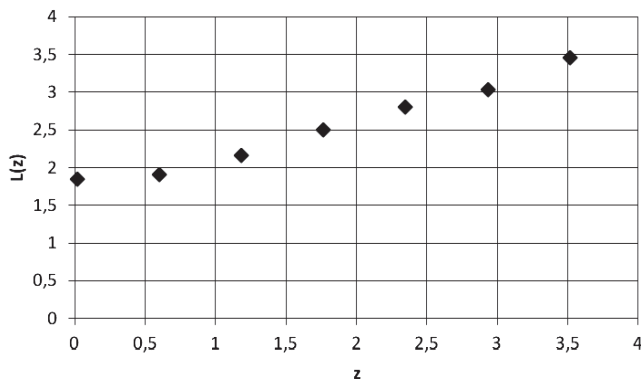
Правдоподобная таблица для $F(x, y, z)$

x	y	z	F	x	y	z	F
0,2733	0,1076	0,0158	1,2218	...	...	...	...
0,8737	0,1076	0,0158	1,4210	0,2733	3,8635	3,5145	5,2687
1,4741	0,1076	0,0158	1,4673	0,8737	3,8635	3,5145	5,4168
2,0745	0,1076	0,0158	2,0474	1,4741	3,8635	3,5145	5,5649
2,6749	0,1076	0,0158	2,6230	2,0745	3,8635	3,5145	5,6150
3,2753	0,1076	0,0158	3,1729	2,6749	3,8635	3,5145	5,6150
3,8757	0,1076	0,0158	3,8406	3,2753	3,8635	3,5145	5,6150
...	...	...	...	3,8757	3,8635	3,5145	5,6150

Таблица 9

Результаты аппроксимации НЛМФА правдоподобной табл/ 8

Разделение переменных на группы в приближающей сумме	$L\{F(x, y, z)\}$	σ – ср. кв. погрешн.	ч.с.с. – число степеней свободы	Приближение принято?
$\{x\}, \{y\}, \{z\}$	$0,2809 \cdot F + 0,0911 \cdot F^2$	1,5	30	Да
$\{x\}, \{y\}$	$0,2809 \cdot F + 0,09109 \cdot F^2$	2,19	36	Нет
$\{x\}, \{z\}$	$0,281 \cdot F + 0,911 \cdot F^2$	2,32	36	Нет
$\{y\}, \{z\}$	$0,281 \cdot F + 0,9109 \cdot F^2$	2,16	36	Нет
$\{x\}, \{y\}, \{z\}$	F	1,71	31	Нет

Рис. 9. Значения функции $L(x)$ Рис. 10. Значения функции $L(y)$ Рис. 11. Значения функции $L(z)$

$$\text{то } F = \frac{-b_1 \pm \sqrt{b_1^2 - 4b_2 L}}{2b_2}.$$

Из рис. 9–11 следует, что $L(x)$, $L(y)$, $L(z)$ отличаются от линейных зависимостей и, скорее всего, они ближе к полиномам 2-й степени. Поэтому, как и в примере 4, функцию F будем искать в виде:

$$F = \sqrt{a_1 x + a_2 x^2 + a_3 y + a_4 y^2 + a_5 z + a_6 z^2 + a_7 + a_8}.$$

Приближение табл. 7 этой формулой со стандартной ошибкой $\sigma \ll 0,001$ дает, формулу $F = \sqrt{x^2 - 0,002x + 0,999y^2 - 0,001y + z^2 + 0,001z - 0,007 + 0,003}$, т.е. практически выявляет истинную закономерность.

Таким образом, НЛМФА возможно применять не только для активного эксперимента, но и для пассивного, что значительно расширяет возможности анализа многомерных данных любого происхождения.

ВЫВОДЫ И ЗАКЛЮЧЕНИЯ

Относительно возможностей применения НЛМФА в практике исследований, проектирования и управления допустимо сделать следующие выводы:

- Возможность с помощью НЛМФА эмпирически выявлять разделение переменных на группы, содержащие не более двух переменных? и выявлять наиболее подходящую шкалу измерения данных сколь угодно большой размерности практически снимает прежде неразрешимую проблему визуального осмысления многомерных данных.
- Возможность с помощью НЛМФА визуализировать приближающие функции от групп разделенных переменных и новую шкалу измерения открывает широкие возможности для создания аналитических эмпирических моделей и формул.
- Возможность с помощью НЛМФА выявлять номографируемые аппроксимирующие модели открывает путь к возрождению номографии и широкому распространению приближенного номографирования.
- НЛМФА открывает новые возможности для планирования и анализа активного эксперимента как в науке, так и в технологиях.
- Применение НЛМФА к неорганизованным наборам данных, к пассивному эксперименту открывает дорогу эмпирическому подходу к областям, в которых преобладает теоретизирование (например, к истории или астрономии).
- Применение НЛМФА к системам из нескольких функций и их преобразований (например, полиномов) обещает выявлять связи между объектами и явлениями, ранее в этих связях не предполагавшиеся.
- Применение НЛМФА «on line» или в автоматическом режиме позволит осуществлять самонастраивающиеся технологические процессы и прогнозирующие системы.

• Прогресс в области вычислительной математики, суперкомпьютеров, создания на их основе искусственного интеллекта при неспособности человека «видеть» в многомерных пространствах грозит опасностью отчуждения его от процесса принятия решений, опасностью потери контроля над процессами, механизмами и системами. Применение НЛМФА способно если не исключить, то существенно уменьшить эту опасность за счет визуализации многомерной информации.

В заключение автор не может не преклониться перед светлой памятью талантливого математика и педагога В.А. Ходакова и рано ушедшего выдающегося ученого и руководителя академика М.Д. Миллионщикова (ученика великого математика А.Н. Колмогорова) за их внимание, критику и моральную поддержку. Нельзя не вспомнить добрым словом и начальника сектора № 6, приверженца курчатовского стиля руководства В.И. Меркина, не препятствовавшего, а скорее поощрявшего инициативы и фантазии своих сотрудников. Автору, инженеру-физику по образованию, без них вряд ли удалось бы не только создать НЛМФА, но и вообще заниматься этой проблемой.

ЛИТЕРАТУРА

1. *Арнольд В.И.* О представлении функций двух переменных в виде $\chi[\varphi(x) + \psi(y)]$ // УМН. – 1957. – Т. 12, № 2. – С. 119–121.
2. *Колмогоров А.Н.* О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения // ДАН СССР. – 1957. – Т. 114, № 5. – С. 953–956.
3. *Мордашев В.М.* Планирование и анализ данных для синтеза многомерных закономерностей (нелинейный многофакторный анализ) // Вопросы атомной науки и техники. Сер. Физика ядерных реакторов. – 2008. – Вып. 2. – С. 3–20.
4. *Мордашев В.М.* О линейной комбинации функций многих переменных, наилучшим образом приближающейся суммой функций меньшего числа переменных // ДАН СССР. – 1971. – Т. 198, № 2. – С. 290.
5. *Рабкин Е.Б., Соколова Е.Г., Лосева Е.И.* Пороги цветоразличения после адаптации к различным уровням насыщенности цветовых полей у аномальных трихроматов // ДАН СССР. – 1976. – Т. 228, № 1. – С. 242–243.
6. *Халтурина Д.А., Коротаев А.В.* Русский крест. – М., 2006.
7. *Землянухин В.В., Кабанов Л.П., Макаровский П.Л. и др.* Скорость смачивания в 7-стержневой сборке из циркониевого сплава при параметрах аварийного охлаждения ВВЭР / Препринт ИАЭ им. И.В. Курчатова, ИАЭ-4118/3, 1985.