Известия ТСХА, выпуск 3, 1983 год

УДК 519.24

ПРИМЕР СЖАТИЯ И ОБРАБОТКИ ИНФОРМАЦИИ МЕТОДОМ РАЗЛОЖЕНИЯ МАТРИЦ ПО ФУНКЦИЯМ УОЛША—АДАМАРА

Н. А. ФРИДЛЕНДЕР

(Кафедра высшей математики)

В предшествующей нашей работе [2] описан метод сжатия информации, представленной большим массивом чисел, и обработки ее с помощью функций Уолша. В данной статье рассматриваемые в [2] методы применены к некоторым конкретным экономическим расчетам. Рассмотрим, например, экономические параметры ряда хозяйств Московской области: пусть f^1 — урожайность зерновых, ц/га; f^2 — уровень интенсивности, тыс. руб. на 1 га; f^3 — среднее число работников на 1 га (1/га)$\cdot$100 %; f^4 — показатель фондоотдачи, руб/руб.; f^5 — нормы минеральных удобрений, ц д. в. на 1 га; f^6 — производительность труда, тыс. руб. на число работников; f^7 — стоимость валовой продукции в расчете на 1 га сельскохозяйственных угодий, тыс. руб/га.

В качестве первого примера возьмем распределение урожайности зерновых в западной части Московской области. За независимые переменные выберем координаты, определяющие местоположение хозяйства.

Впишем в западную часть области квадрат, разделенный на 16 малых квадратов. Местоположение квадратов определяется тремя опорными точками, за которые мы принимали 3 города: Солнечногорск — элемент a_{14} матрицы, Шаховская — a_{11}, Нарофоминск — a_{43}. Элементами построенной матрицы I 4$\times$4 являются средние урожайности зерновых для каждого из малых квадратов, а матрицы II — коэффициенты разложения функции (коэффициенты Фурье) по 16 функциям Уолша — Адамара. Алгоритм преобразования Уолша — Адамара подробно описан в [2] и мы его здесь не приводим.

$$\begin{vmatrix} 23 & 24 & 25 & 24 \\ 13 & 18 & 19 & 28 \\ 14 & 22 & 25 & 23 \\ 18 & 21 & 20 & 20 \end{vmatrix} \qquad \begin{vmatrix} 337 & 31 & -11 & -23 \\ 11 & -5 & 15 & -5 \\ 18 & 25 & 1 & 17 \\ 23 & 3 & -13 & 11 \end{vmatrix} : 4$$

I II

Матрицы III и IV дают соответственно урожайности зерновых для всей Московской области и коэффициенты Фурье после сжатия.

$$f^1_{ij} = \begin{vmatrix} 22 & 18{,}7 & 34 & 19{,}5 & 26{,}2 & 28 & 28{,}6 & 28{,}9 \\ 16{,}7 & 24{,}6 & 15{,}7 & 23{,}0 & 24{,}0 & 24{,}3 & 27{,}3 & 32{,}3 \\ 19{,}1 & 23{,}4 & 20{,}3 & 19{,}8 & 36{,}0 & 27{,}8 & 24{,}6 & 30{,}5 \\ 14{,}6 & 25{,}6 & 20{,}3 & 24{,}2 & 25{,}8 & 28{,}8 & 40{,}0 & 41{,}1 \\ 11{,}2 & 23{,}9 & 14{,}5 & 23{,}4 & 26{,}9 & 27{,}7 & 23{,}4 & 42{,}0 \\ 19{,}6 & 26{,}5 & 24{,}6 & 20{,}6 & 25{,}8 & 28{,}5 & 27{,}2 & 27{,}3 \\ 21{,}6 & 24{,}4 & 28{,}9 & 27{,}1 & 24{,}2 & 25{,}4 & 31{,}6 & 27{,}4 \\ 35{,}1 & 48{,}9 & 30{,}7 & 24{,}7 & 36{,}7 & 21{,}7 & 32{,}7 & 23{,}4 \end{vmatrix} \qquad \alpha^1 = \begin{vmatrix} 370{,}9 & -92{,}3 & 46{,}3 & 21{,}1 \\ -2{,}9 & -14{,}1 & 3{,}7 & -3{,}1 \\ 3{,}5 & 23{,}5 & -6{,}3 & -2{,}7 \\ -17{,}1 & -27 & 0{,}3 & -0{,}1 \end{vmatrix} : 4$$

III IV

Анализируя элементы матрицы III, видим, что урожайность в хозяйствах меняется от 11 до 42, а коэффициенты спектра претерпевают весьма сильные изменения — от 0,1/4 до 370,9/4, что и дает возможность произвести сжатие информации, отбрасывая малые по абсолютной величине члены. Коэффициенты Фурье оценивают матрицу «в целом», выделяя незаметные на первый взгляд закономерности распределения в ней чисел. Из матриц коэффициентов Фурье (II и IV) можно просто получить среднюю урожайность зерновых: для западной зоны Московской области она равна

$$\alpha^1_{11}/4 = (337/16) = 21{,}06 \text{ ц/га}, \qquad (1)$$

а для всей Московской области — 23,18 ц/га.

Матрицами V и VI представлены соответственно значения уровней интенсивности хозяйств, т. е. затраты, отнесенные к площади сельскохозяйственных угодий, по всей Московской области, и коэффициенты Фурье после сжатия.

$$f^2_{ij} = \begin{vmatrix} 0{,}38 & 0{,}53 & 0{,}61 & 0{,}62 & 0{,}7 & 0{,}75 & 0{,}98 & 1{,}22 \\ 0{,}425 & 0{,}58 & 0{,}59 & 0{,}66 & 0{,}69 & 0{,}81 & 0{,}94 & 0{,}99 \\ 0{,}48 & 0{,}57 & 0{,}62 & 0{,}66 & 0{,}72 & 0{,}8 & 0{,}95 & 1{,}17 \\ 0{,}51 & 0{,}58 & 0{,}58 & 0{,}65 & 0{,}71 & 0{,}74 & 0{,}87 & 0{,}98 \\ 0{,}51 & 0{,}54 & 0{,}6 & 0{,}66 & 0{,}43 & 0{,}74 & 0{,}84 & 1{,}61 \\ 0{,}48 & 0{,}59 & 0{,}59 & 0{,}64 & 0{,}74 & 0{,}75 & 0{,}85 & 1{,}06 \\ 0{,}45 & 0{,}57 & 0{,}63 & 0{,}64 & 0{,}68 & 0{,}81 & 0{,}82 & 1{,}04 \\ 0{,}51 & 0{,}52 & 0{,}61 & 0{,}66 & 0{,}68 & 0{,}77 & 0{,}88 & 1{,}28 \end{vmatrix} \qquad \alpha^2 = \begin{vmatrix} 11{,}54 & -2{,}46 & 0{,}78 & -1{,}7 \\ 0 & 0 & -0{,}16 & 0{,}16 \\ 0{,}05 & -0{,}17 & -0{,}11 & -0{,}001 \\ -0{,}11 & -0{,}05 & 0{,}09 & -0{,}17 \end{vmatrix} : 4$$

V VI

Среднее значение уровня интенсивности по Московской области равно

$\alpha^2_{11}/4=11{,}54/16=0{,}721$ тыс. руб/га. (2)

Среднее значение фондоотдачи f^4, т. е. отношение стоимости валовой продукции к стоимости основных фондов, для области (матрица VII) равно

$\alpha^4_{11}/4=6{,}72/16=0{,}42$ руб/руб. (3)

Соответствующие коэффициенты Фурье после сжатия показаны на матрице VIII.

$$f^4_{ij}=\begin{vmatrix} 0{,}32 & 0{,}34 & 0{,}25 & 0{,}27 & 0{,}35 & 0{,}38 & 0{,}43 & 0{,}27 \\ 0{,}33 & 0{,}37 & 0{,}39 & 0{,}46 & 0{,}36 & 0{,}42 & 0{,}46 & 0{,}34 \\ 0{,}4 & 0{,}43 & 0{,}35 & 0{,}33 & 0{,}29 & 0{,}36 & 0{,}45 & 0{,}36 \\ 0{,}35 & 0{,}43 & 0{,}3 & 0{,}46 & 0{,}29 & 0{,}44 & 0{,}5 & 0{,}54 \\ 0{,}38 & 0{,}34 & 0{,}36 & 0{,}4 & 0{,}46 & 0{,}53 & 0{,}55 & 0{,}21 \\ 0{,}34 & 0{,}53 & 0{,}42 & 0{,}29 & 0{,}34 & 0{,}49 & 0{,}6 & 0{,}5 \\ 0{,}48 & 0{,}38 & 0{,}49 & 0{,}41 & 0{,}64 & 0{,}37 & 0{,}73 & 0{,}4 \\ 0{,}51 & 0{,}7 & 0{,}62 & 0{,}63 & 0{,}49 & 0{,}41 & 0{,}34 & 0{,}43 \end{vmatrix}$$

VII

$$\alpha^4=\begin{vmatrix} 6{,}72 & -0{,}12 & 0{,}14 & -0{,}1 \\ -0{,}68 & -0{,}04 & 0{,}1 & -0{,}1 \\ 0{,}24 & 0{,}24 & -0{,}26 & 0{,}02 \\ -0{,}41 & -0{,}24 & -0{,}14 & 0{,}1 \end{vmatrix}:4$$

VIII

Матрицы III, V и VII получены способом систематизирования параметров, описанным ниже, и последующим сжатием.

Вычислим теперь среднеквадратическое отклонение урожайности зерновых по всей области

$$\sigma(f^1)=\sqrt{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(f^1_{ij})^2-[M(f^1)]^2}=$$

$$=\frac{1}{n}\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha^1_{ij})^2-(\alpha^1_{11})^2}=6{,}92 \text{ ц/га}, \quad (4)$$

здесь $M[f^1]$ — математическое ожидание функции f^1.

Аналогично получаем

$\sigma(f^2)=0{,}195$, ц/га. (5)

Если при вычислении $\sigma(f^1)$ оставить только 4 наибольших коэффициента, превышающих по абсолютному значению 20/4, то получим

$\sigma(f^1)\approx 6{,}74$, ц/га, (6)

т. е. погрешность выражается в десятых долях.

Найдем далее коэффициент корреляции между f^1 и f^2, т. е. между урожайностью и уровнем интенсивности. Согласно [2] корреляционный момент может быть вычислен через коэффициенты Фурье по формуле

$$K(f^1f^2)=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha^1_{ij}\alpha^2_{ij}-\frac{\alpha^1_{11}\alpha^2_{11}}{n^2}=0{,}8795, \quad (7)$$

где α^1_{11} и α^2_{11} — первые коэффициенты разложения соответственно функций f^1 и f^2.

Коэффициент корреляции между f^1 и f^2 равен 1:

$$\bar{r}^{12}=\frac{K(f^1f^2)}{\sigma(f^1)\,\sigma(f^2)}=\frac{0{,}8795}{6{,}92\cdot 0{,}195}=0{,}653. \quad (8)$$

Если в формулах (4), (5), (7) оставить под знаком суммы только четыре наибольших слагаемых, то $r^{12}=0{,}665$, т. е. имеет погрешность в сотых долях.

Аналогичные расчеты дают

$r^{14}=0{,}316$; $r^{24}=0{,}1848$. (9)

Проведенные расчеты показали, что при вычислении коэффициентов корреляции можно производить сжатие информации в 4—5 раз, что мало сказывается на точности результата.

Как уже указывалось, матрица I

строилась методом вписывания квадрата в западную часть области. При этом часть хозяйств, расположенных вблизи границы области, может не войти в исследование. Поэтому применим другой метод — метод криволинейных координат. На рисунке приведены контуры Московской области. Вся область разделена криволинейной сеткой на 64 подобласти. Желательно, чтобы все подобласти были равновеликими, но это не очень существенно для результатов исследования. Площадь г. Москвы из рассмотрения исключалась. В каждой из подобластей на рисунке записано значение средней урожайности зерновых, соответствующее этой подобласти. Таким образом были получены квадратная матрица IX 8×8 для функции f^1 и матрица соответствующих коэффициентов разложения X.

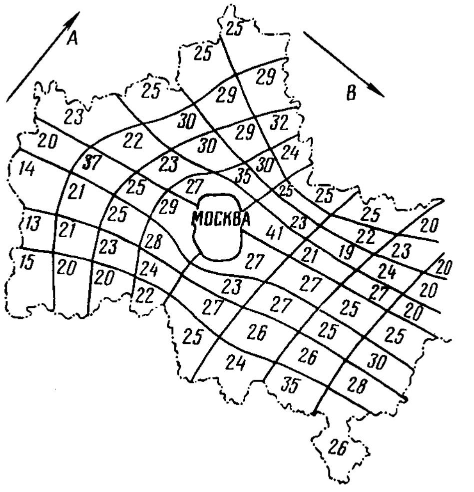

Схема деления Московской области сеткой криволинейных координат.

$$
\begin{array}{cc}
 & \overrightarrow{\;B\;} \\
A\uparrow & \left|\begin{array}{cccccccc}
25 & 29 & 32 & 24 & 25 & 25 & 20 & 20 \\
25 & 29 & 29 & 30 & 25 & 22 & 23 & 20 \\
25 & 30 & 30 & 35 & 23 & 19 & 24 & 20 \\
23 & 22 & 23 & 27 & 41 & 21 & 27 & 20 \\
20 & 21 & 25 & 29 & 27 & 27 & 25 & 25 \\
14 & 21 & 25 & 28 & 23 & 27 & 25 & 30 \\
13 & 21 & 23 & 24 & 27 & 26 & 26 & 28 \\
15 & 20 & 20 & 22 & 25 & 24 & 35 & 26
\end{array}\right|
\end{array}
$$

IX

$$
\left|\begin{array}{cccccccc}
1584 & -26 & -86 & -60 & -12 & -30 & -90 & 0 \\
80 & 114 & -42 & 100 & -20 & 22 & 10 & 24 \\
-26 & -8 & 28 & 10 & -46 & -12 & 24 & -2 \\
8 & 58 & -18 & 12 & -16 & -6 & 62 & -36 \\
0 & -50 & -2 & 40 & 50 & -16 & -10 & 56 \\
-10 & -52 & -28 & 22 & -4 & -54 & 2 & -12 \\
-8 & 30 & 30 & -32 & -50 & 6 & 22 & -24 \\
6 & 60 & -12 & 6 & -14 & -8 & 32 & -22
\end{array}\right| :8
$$

X

Рассмотрим наиболее характерные коэффициенты Фурье. Коэффициент α_{11} определяет среднее значение урожайности зерновых по всей Московской области

$$M(f^1) = \alpha_{11}/8 = 1584/64 = \\ = 24{,}75,\ \text{ц/га}. \qquad (10)$$

Коэффициент $\alpha_{21}/8 = 1{,}25$ ц/га показывает, что в северо-восточной части Московской области урожайность зерновых на 1,25 ц/га больше среднего значения, а на юго-западе — меньше на ту же величину. Элемент $\alpha_{12}/8 = 26/64 = 0{,}41$ ц/га свидетельствует, что на северо-западе урожайность в среднем несколько ниже, чем на юго-востоке (направление B). Коэффициент $\alpha_{22}/8 = 114/64 = 1{,}78$ ц/га указывает на то, что в среднем на направлении главной диагонали матрицы IX урожайность на 1,78 ц/га больше среднего значения, а на направлении побочной диагонали на такую же величину меньше среднего значения.

Из матрицы X следует, что более двух третей всех ее элементов меньше по абсолютному значению 30/8 и их можно отбросить, как это мы делали раньше. Для достижения большей точности расчетов и большего сжатия информации рассматриваемую область нужно было бы разбить криволинейной сеткой не на 8×8, а на большее число подобластей, например на 16×16.

До сих пор мы рассматривали изменение параметров $f^1, f^2, \ldots, f^7$ в зависимости от расположения хозяйств, которое характеризовалось координатами x и y. Очевидно, информация о хозяйствах области не исчерпывается этими функциями, их может быть и больше. Вообще говоря, каждое хозяйство области характеризуется вектором $F_к$ $(x;\ y;\ f^1,\ f^2, \ldots, f^m)$, где $к$ — номер хозяйства; x, y — его местоположение, а $f^1, f^2, \ldots, f^m$ — производственные параметры. Информация о работе сельскохозяйственных предприятий представляется в виде системы векторов $F_к$. Если при обработке информации ее не систематизировать, то сжатие окажется не очень эффективным.

Сжатие будет тем больше, чем сильнее зависят исследуемые функции от систематизированных (независимых) переменных. При этом обычно в матрице коэффициентов Фурье наибольшие коэффициенты располагаются в левом верхнем углу. Если, например, обрабатывается информация об уровне интенсивности в области, то для большего сжатия следует расположить хозяйства в порядке возрастания обеспеченности фондами или рабочей силой; если обрабатывается информация о стоимости валовой продукции, то за систематизированные переменные можно взять обеспеченность рабочей силой, количество удобрений или качество почвы. Если за независимую переменную принять параметр f^i, то нужно вначале расположить хозяйства в порядке возрастания его значения. При этом получится простая статистическая таблица. При выборе двух независимых переменных, например f^i и $f^\varkappa$, при наличии 256 хозяйств (как в нашем случае) нужно простую статистическую таблицу представить в виде интервальной таблицы по параметру f^i (в данном примере 16 интервалов) и в каждом из интервалов расположить хозяйства в порядке возрастания параметра $f^\varkappa$. В результате получим корреляционную таблицу.

Можно выбрать и три (или более) независимые переменные при достаточно большом числе хозяйств (более 4000), но получаемые 3-мерные (и т. д.) матрицы менее удобны, чем квадратные.

Мы исследовали функции $f^\varkappa$ ($k=1, 2, ..., 7$). За независимые переменные взяли f^2 и f^6. Были построены матрицы 16×16 по возрастанию параметров f^2 и f^6 соответственно по горизонтальной и вертикальной осям (т. е. соответственно по столбцам и по строкам). Из-за громоздкости матриц мы их не приводим. Эти матрицы подвергали двумерному сжатию методом Уолша — Адамара. Для всех функций были рассчитаны математические ожидания $M(f^i)$ и средние квадратические отклонения $\sigma(f^i)$ (табл. 1).

Были рассчитаны также взаимные коэффициенты корреляции — всего 21 коэффициент (табл. 2). Расчет их, проведенный традиционными методами, потребовал бы значительно большей работы.

Таблица 1

Математическое ожидание и среднее квадратическое отклонение рассматриваемых функций

Показатель	f^1	f^2	f^3	f^4	f^5	f^6	f^7
$M[f^i]$	23,181	0,721	9,887	0,42	2,7	5,919	0,5825
$\sigma[f^i]$	6,917	0,1946	2,682	0,0622	0,633	4,829	0,192

Таблица 2

Коэффициенты корреляции

Показатель	f^1	f^2	f^3	f^4	f^5	f^6	f^7
f^1	1	0,653	0,6092	0,3167	0,6983	0,1349	0,6855
f^2		1	0,918	0,1848	0,6259	0,06989	0,9512
f^3			1	—0,035	0,6503	—0,0654	0,8626
f^4				1	—0,0127	0,4857	0,3201
f^5					1	—0,0797	0,6255
f^6						1	0,192
f^7							1

Из табл. 2 видно, что более тесные корреляционные связи существуют между уровнем интенсивности и числом работников, приходящихся на 1 га: $r^{23}=0,918$; между уровнем интенсивности и стоимостью валовой продукции, отнесенной к площади сельскохозяйственных угодий: $r^{27}=0,9512$; между числом работников, приходящихся на 1 га и стоимостью валовой продукции, отнесенной к площади сельскохозяйственных угодий: $r^{37}=0,8626$.

Значения r^{12}, r^{13}, r^{15}, r^{17}, r^{25}, r^{35}, r^{57} заключены в пределах 0,5—0,7. Остальные коэффициенты корреляции весьма малы, что указывает на отсутствие линейной зависимости между параметрами.

Автор приносит большую благодарность доценту кафедры статистики В. С. Филимонову за предоставленные статистические сведения по хозяйствам Московской области и ряд ценных обсуждений.

Выводы

1. В работе приведены примеры обработки технико-экономических показателей работы хозяйств Московской области методом отображения Уолша — Адамара.

2. Показано удобство предложенного метода при определении средних значений показателей для хозяйств области, среднеквадратических отклонений, коэффициентов корреляции.

3. Предлагаемый метод особенно выгоден при обработке данных большого числа предприятий.

4. Показана возможность сжатия информации в 4—5 раз, что важно при ее хранении и обработке.

5. Объем вычислений технико-экономических показателей хозяйств области уменьшается пропорционально сжатию имеющейся информации.

6. Коэффициенты Фурье при низкочастотных функциях Уолша дают возможность еще до вычисления основных характеристик функций оценить их поведение в целом по области.

ЛИТЕРАТУРА

1. Вентцель Е. С. Теория вероятностей. М.: Наука, 1969. — 2. Фридлендер Н. А. Сжатие информации с больших областей и обработка ее с помощью функций Уолша. — Изв. ТСХА, вып. 3, 1982, с. 179—188. — 3. Обработка изображений и цифровая фильтрация. / Под ред. Т. Хуанга. М.: Мир, 1979.

Статья поступила 2 февраля 1983 г.

SUMMARY

On the basis of building up multiplied system of stepped functions of Walsh of two measured region and finding of reflections of Walsh-Adamar it was shown the practical possibility of usage of these methods for approximation of these production functions with set accuracy and conciseness of information on the work of agricultural enterprises in large regions.

The example of treatment of agricultural enterprises' indices in Moscovskaja district was given. The calculation of correlation dependence between parameters on statistical data with the help of Furje coefficients in the field of reflection was conducted.