А. Б. Хмельницкая, Уточнение понятия решения в задаче инвариантного шкалирования, *Автомат. и телемех.*, 1989, выпуск 6, 123–128

УДК 517.977

УТОЧНЕНИЕ ПОНЯТИЯ РЕШЕНИЯ В ЗАДАЧЕ ИНВАРИАНТНОГО ШКАЛИРОВАНИЯ

ХМЕЛЬНИЦКАЯ А. Б.

(Ленинград)

Приводится условие корректности выбора в качестве функции шкалирования наиболее глубокой стационарной точки рассматриваемого функционала адекватности структур близости между объектами при переходе от исходного описания к агрегированному.

1. Введение

Одним из возможных подходов к построению агрегированного описания больших массивов эмпирической информации является использование методов многомерного инвариантного шкалирования [1—4]. При этом под агрегированием понимается построение такого отображения t из исходного признакового пространства X на шкалу укрупненного описания — евклидово пространство R^k заданной размерности $k\geqslant 1$, при котором структуры близости между объектами при переходе от исходного описания к агрегированному максимально согласованы в смысле выбранного критерия — фиксированного функционала. В качестве критерия согласованности структур близости при отображении f: $X\to R^k$ может рассматриваться функционал

$$\Psi(f)=\int_X\int_X\left[r(x,y)-\sum_{i=1}^k[f_i(x)-f_i(y)]^2\right]^2 d\mu(x)\,d\mu(y),$$

где μ и $r(x, y)$ — соответственно вероятностная мера и функция близости в X (близость образов исходных объектов в R^k оценивается посредством квадрата евклидовой метрики). Тогда задача инвариантного k-мерного шкалирования ставится как задача поиска глобального минимума функционала $\Psi(f)$ на всей его области определения $L^4(X, \mu; R^k)$. Известный метод построения отображения t, доставляющего глобальный минимум функционалу $\Psi(f)$ и называемого функцией шкалирования, основан на поиске стационарных точек функционала $\Psi(f)$ в $L^4(X, \mu; R^k)$. При этом из множества всех стационарных точек выбирается наиболее глубокая, т. е. доставляющая $\Psi(f)$ наименьшее значение. Этот способ определения значений функции шкалирования t был бы полностью оправдан, если бы удалось показать, что функционал $\Psi(f)$ всегда достигает своего глобального минимума. Однако это утверждение в общем случае не доказано [1]. Существование глобального минимума функционала $\Psi(f)$ без каких-либо ограничений на f установлено только для случая дискретной меры μ. Для непрерывных мер μ факт существования глобального минимума доказан лишь при выполнении ряда ограничений, в том числе весьма жестких ограничений на искомую функцию шкалирования. В связи со сказанным приходится несколько видоизменить постановку задачи шкалирования и под функцией шкалирования вместо глобального минимума функционала $\Psi(f)$ понимать его локальный минимум, причем при наличии нескольких локальных минимумов выбирать наиболее глубокий. В таком случае воз-

[1] Имеющиеся по этому поводу результаты изложены в [4, п. 3.3.].

можность выбора наиболее глубокой стационарной точки функционала $\Psi(f)$ в качестве решения задачи шкалирования требует согласования с необходимыми условиями локального минимума этого функционала. Исследованию этого вопроса и посвящена настоящая статья.

Поясним подробнее сказанное выше. Известно [3], что при построении k-мерной вектор-функции шкалирования вместо всего пространства $L^4(X, \mu; R^k)$ достаточно ограничиться рассмотрением функций класса [2]

$$L_k=\{f\in L^4(X;\ \mu;\ R^k)\,|\,\bar{f}=0;\ \ (f_i,\ f_j)=0,\ \ i,\ j\in\overline{1,\ k},\ i\neq j\}.$$

Если глобальный минимум функционала $\Psi(f)$ на функциях класса L_k достигается, т. е. решение k-мерной задачи шкалирования $t=\operatorname*{argmin}_{f\in L_k}\Psi(f)$ существует, то нахождение функции шкалирования t может быть сведено к поиску стационарных точек этого функционала. Относительно одномерных составляющих вектор-функции $t\in L_k$, являющейся стационарной точкой функционала $\Psi(f)$ в банаховом пространстве $L^4(X,\ \mu;\ R^k)$ (в частности, относительно координатных функций вектор-функции шкалирования), выписана система интегро-степенных уравнений третьей степени [2], названная системой уравнений шкалирования: при почти всех $x\in\in(X,\ \mu)$ для всех $i\in\overline{1,\ k}$

$$(1)\qquad t_i(x)\Big[\sum_{j=1}^{k}t_j^2(x)+b_i(t;x)\Big]-e_i(t;x)=0,$$

где для любых $f\in L_k$, $x\in X$, $i\in\overline{1,\ k}$

$$(2)\qquad b_i(f;x)=\int_X\sum_{j=1}^{k}f_j^2d\mu+2\int_X f_i^2d\mu-\int_X r(x,y)\,d\mu(y),$$

$$e_i(f;x)=\int_X f_i\sum_{j=1}^{k}f_j^2d\mu-\int_X r(x,y)\,f_i(y)\,d\mu(y).$$

Показано [3, 4], что решение $t(x)$ системы уравнений шкалирования (1), т. е. стационарная точка функционала $\Psi(f)$, почти везде в $(X,\ \mu)$ однозначно определяется величинами $b_i(t;\ x)$, $e_i(t;\ x)$, $i\in\overline{1,\ k}$ и значением характеристической функции шкалирования

$$(3)\qquad u_t(x)=\sum_{i=1}^{k}t_i^2(x),$$

которое, в свою очередь, является корнем характеристического уравнения

$$(4)\qquad \varphi(u)=0.$$

Здесь функция $\varphi(u)$ задается равенством

$$(5)\qquad \varphi(u)=u-\sum_{i\in I(t;\,x)}e_i^2(t;x)/[u+b_i(t;x)]^2,$$

[2] Здесь через $\bar{f}$ обозначен вектор средних значений f, т. е. $\bar{f}=(\bar{f}_1,\ldots,\bar{f}_k)$, $\bar{f}_i=\int_X f_i\,d\mu$, а через $(f_i,\ f_j)$ – скалярное произведение в смысле $L^2(X,\ \mu)$ координатных составляющих $f_i(x)$ и $f_j(x)$, т. е. $(f_i,\ f_j)=\int_X f_if_j\,d\mu$.

а множество индексов $I(f;\ x)$ для любого $f \in L_k$ определяется соотношением

$$I(f;\ x)=\{i \in \overline{1,\ k} \mid e_i(f;\ x) \neq 0\}. \tag{6}$$

Исходя из определения функции шкалирования t как точки глобального минимума функционала $\Psi(f)$ и в предположении о том, что глобальный минимум функционала $\Psi(f)$ на L_k достигается, t является наиболее глубокой стационарной точкой этого функционала, т. е. на функции t из множества всех решений системы уравнений шкалирования (1) функционал $\Psi(f)$ принимает наименьшее значение. В [4, 5] установлено, что для решения $t \in L_k$ системы уравнений шкалирования (1), на котором достигается наименьшее значение функционала $\Psi(f)$ на множестве всех решений системы (1), почти везде в $(X,\ \mu)$ соответствующее значение характеристической функции $u_t(x)$ совпадает с максимальным корнем $\hat{u}_t(x)$ характеристического уравнения (4) в точке x. Однако в связи с тем, что существование глобального минимума функционала $\Psi(f)$ на функциях класса L_k в общем случае не доказано, приходится ослабить понятие решения в задаче k-мерного шкалирования. Поэтому в дальнейшем под функцией шкалирования будет пониматься функция класса L_k, на которой достигается локальный минимум функционала $\Psi(f)$ в смысле пространства $L^4(X,\ \mu;\ R^k)$. При этом при наличии нескольких локальных минимумов в качестве функции шкалирования будет выбираться функция, доставляющая $\Psi(f)$ наиболее глубокий локальный минимум (если таковой существует). Тогда для возможности использования указанного способа построения функции шкалирования, как наиболее глубокой стационарной точки функционала $\Psi(f)$, по крайней мере надо, чтобы в наиболее глубокой стационарной точке выполнялись необходимые условия локального минимума [3].

В данной работе выводится система неравенств, которой должны удовлетворять координатные составляющие k-мерной функции шкалирования вследствие ее определения как точки локального минимума функционала $\Psi(f)$. А затем устанавливается условие, при выполнении которого выбор наиболее глубокой стационарной точки функционала $\Psi(f)$ в качестве функции шкалирования корректен с точки зрения уточненного определения решения задачи шкалирования. При этом корректность понимается в том смысле, что справедливость этого условия обеспечивает выполнение в наиболее глубокой стационарной точке указанной системы неравенств.

2. Необходимое условие локального минимума

Исходя из определения функции шкалирования как точки локального минимума функционала $\Psi(f)$, верна

Теорема 1. Если $t \in L_k$ — k-мерная функция шкалирования, то для почти всех $x \in (X,\ \mu)$ при всех $i \in \overline{1,\ k}$ выполняются неравенства

$$u_t(x)+2t_i^2(x)+b_i(t;\ x) \geqslant 0, \tag{7}$$

где $u_t(x)$ — характеристическая функция шкалирования, определяемая равенством (3), а величины $b_i(t;\ x)$ задаются соотношениями (2).

---

[3] Если глобальный минимум не достигается, то наиболее глубокая стационарная точка, вообще говоря, не обязана быть точкой локального минимума. Существует ли реально для данной задачи такая ситуация, когда глобальный минимум не достигается, пока неизвестно.

(Доказательство приведено в приложении.)

Отметим, что проверка справедливости системы неравенств (7) является существенным моментом при построении функции шкалирования на основе решения системы уравнений шкалирования (1).

Так, например, вектор-функция $f \in L_k$, тождественно равная нулю по мере μ, т. е. для почти всех $x \in (X, \mu)$ $f(x)=0$, для любой k-мерной задачи, задаваемой тройкой (X, μ, r), удовлетворяет системе уравнений шкалирования (1). Однако при почти всех $x \in (X, \mu)$ для всех $i \in \overline{1, k}$

$$u_f(x) + 2f_i^2(x) + b_i(f; x) = -\int_X r(x, y)\, d\mu(y) < 0.$$

Тем самым неравенства (7) никогда не выполняются, и вследствие теоремы 1 функция $f\colon X \to R^k$, задаваемая для почти всех $x \in (X, \mu)$ равенством $f(x)=0$, никогда не является функцией шкалирования.

Таким образом, справедлива

Теорема 2[4]. Функция шкалирования всегда отлична от константы[5].

3. Условие корректности выбора наиболее глубокой стационарной точки функционала $\Psi(f)$ в качестве функции шкалирования

Как уже отмечалось выше, утверждение о совпадении почти везде в (X, μ) значений характеристической функции шкалирования $u_t(x)$ с максимальным корнем $\hat{u}_t(x)$ характеристического уравнения (4) в точке x, т. е. равенство

$$u_t(x) = \hat{u}_t(x) \text{ для почти всех } x \in (X, \mu), \tag{8}$$

установлено в предположении о том, что функция шкалирования совпадает с наиболее глубокой стационарной точкой функционала $\Psi(f)$. Тогда в свете приведенного выше уточненного определения функции шкалирования корректность выбора максимального корня характеристического уравнения в качестве значения характеристической функции шкалирования требует, по крайней мере, чтобы наиболее глубокая стационарная точка функционала $\Psi(f)$ удовлетворяла необходимым условиям локального минимума.

Введем обозначения. Для любых $f \in L_k$ и $x \in X$ положим

$$B(f; x) = \max_{i \in I(f; x)} \{-b_i(f; x)\}, \quad \overline{B}(f; x) = \max_{\substack{i \in \overline{1, k} \\ i \overline{\in} I(f; x)}} \{-b_i(f; x)\}.$$

где величины $b_i(f; x)$ и $I(f; x)$ определяются посредством (2) и (6) соответственно.

Имеет место

Теорема 3. (Условие корректности.) Пусть $t \in L_k$ — k-мерная функция шкалирования. Тогда для того, чтобы при почти всех $x \in (X, \mu)$ выполнение равенства (8) не противоречило системе неравенств (7), необходимо и достаточно, чтобы для почти всех $x \in (X, \mu)$ из

$$\overline{B}(t; x) > B(t; x) \tag{9}$$

[4] Для одномерного случая это утверждение впервые было сформулировано в [1]. Однако в приведенном там доказательстве, основанном на иных соображениях, допущена неточность.

[5] Равенство $t = \text{const} (\mathrm{mod}\, \mu)$ при условии $t \in L_k$ равносильно равенству $t = 0\ (\mathrm{mod}\, \mu)$.

следовало

(10) $\varphi(\overline{B}(t;\ x))\leqslant 0,$

где функция φ задается соотношением (5).

(Доказательство приведено в приложении.)

Следствие. Пусть функция шкалирования $t\in L_k$ такова, что при почти всех $x\in(X,\ \mu)$ для всех $i\in\overline{1,\ k}$ $e_i(t;\ x)\neq 0$, т. е. при почти всех $x\in(X,\ \mu)$ $I(t,\ x)=\{i\in\overline{1,\ k}\}$ [6]. Тогда при выборе во всех точках $x\in X$ максимального корня характеристического уравнения (4) в качестве значения характеристической функции $u_t(x)$ необходимое условие локального минимума функционала $\Psi(f)$, выраженное системой неравенств (7), на функции t оказывается выполненным.

Отметим, что при построении функции шкалирования на основе решения системы уравнений шкалирования (1) с учетом равенства (8) проверка свойства функции шкалирования удовлетворять системе неравенств (7) может быть сведена к проверке более удобного условия (9)–(10).

В заключение сделаем следующее замечание. Как уже упоминалось выше, в случае решения задачи шкалирования на конечной выборке функционал $\Psi(f)$ достигает на некоторой функции $t\in L_k$ своего глобального минимума, который в таком случае естественным образом совпадает с его наиболее глубокой стационарной точкой (тем самым условие (9)–(10) на функции t автоматически оказывается выполненным). Тем не менее, так как построение функции шкалирования сводится к нахождению конечного набора порождающих ее параметров, а для определения системы параметров шкалирования используются градиентные методы, то для того, чтобы исключить возможность остановки локального градиентного метода в точке перегиба, целесообразно, прежде чем объявить точку остановки градиентного спуска решением задачи шкалирования, проверить выполнение в этой точке условия (9)–(10) (соответственно переписанного в терминах системы параметров шкалирования).

ПРИЛОЖЕНИЕ

Доказательство теоремы 1. Нетрудно показать, что второй дифференциал функционала $\Psi(f)$ в точке t для любого $h\in L^4(X,\ \mu;\ R^k)$ имеет вид

$$d^2\Psi(t)=4\iint\limits_{XX}\left\{\left[\sum_{i=1}^{k}[t_i(x)-t_i(y)]^2-r(x,\ y)\right]\sum_{i=1}^{k}[h_i(x)-h_i(y)]^2+\right.$$

$$\left.+2\left[\sum_{i=1}^{k}[t_i(x)-t_i(y)][h_i(x)-h_i(y)]\right]^2\right\}d\mu(x)\,d\mu(y).$$

В силу свойств локального минимума функционала в банаховом пространстве [6, гл. 10, § 2] для любого $h\in L^4(X,\ \mu;\ R^k)$ $d^2\Psi(t)\geqslant 0$. Последнее неравенство, в частности, имеет место для любого $h\in L^4(X,\ \mu;\ R^k)$, такого, что лишь при одном $i\in\overline{1,\ k}$ $h_i(x)$ отлично от тождественного нуля по мере μ, а для всех $j\in\overline{1,\ k}$, $j\neq i$ при почти всех $x\in(X,\ \mu)$ $h_j(x)=0$. Следовательно, так как всегда

$$[h_i(x)-h_i(y)]^2\geqslant 0,$$

то для всех $i\in\overline{1,\ k}$ при почти всех $x,\ y\in(X,\ \mu)$ справедливо неравенство

$$\sum_{j=1}^{k}[t_j(x)-t_j(y)]^2-r(x,\ y)+2[t_i(x)-t_i(y)]^2\geqslant 0$$

---

[6] Это условие, в частности, означает, что мера множества особых точек функции t, т. е. точек $x\in X$, в которых хотя бы при одном $i\in\overline{1,\ k}$ $u_t(x)+b_i(t;\ x)=0$, равна нулю.

или, что то же самое,

$$\sum_{j=1}^{k} t_j^2(x)+\sum_{j=1}^{k} t_j^2(y)-2\sum_{j=1}^{k} t_j(x)t_j(y)+2t_i^2(x)+2t_i^2(y)-$$
$$-4t_i(x)t_i(y)-r(x,y)\geqslant 0.$$

Проинтегрируем полученное неравенство по y и воспользуемся равенствами

$$\int\limits_X t_i d\mu=0,$$

которые в силу условия $t\in L_k$ имеют место при всех $i\in\overline{1,\ k}$. Тогда для всех $i\in\overline{1,\ k}$ при почти всех $x\in(X,\ \mu)$

$$\sum_{j=1}^{k} t_j^2(x)+2t_i^2(x)+\int\limits_X\sum_{j=1}^{k} t_j^2 d\mu+2\int\limits_X t_i^2 d\mu-\int\limits_X r(x,y)\,d\mu(y)\geqslant 0,$$

Отсюда, используя обозначения (2) и определение характеристической функции (3), устанавливаем справедливость теоремы.

Доказательство теоремы 3. Легко видеть, что вследствие равенства (8) и теоремы 1 [5] для почти всех $x\in(X,\ \mu)$

$$u_t(x)>B(t;\ x),$$

т. е. для почти всех $x\in(X,\ \mu)$ при всех $i\in I(t;\ x)$ неравенство (7) оказывается выполненным.

Далее, для почти всех $x\in(X,\ \mu)$ из условия $i\in\overline{1,\ k}$, $i\equiv I(t;\ x)$, т. е. $l_i(t;\ x)=0$, в силу теорем 3 [3] и 2.5.8 [4] следует, что $t_i(x)=0$.

Следовательно, при почти всех $x\in(X,\ \mu)$, таких, что $I(t;\ x)\neq\{i\in\overline{1,\ k}\}$ для всех $i\in\overline{1,\ k}$, $i\equiv I(t;\ x)$, неравенство (7) равносильно неравенству $u_t(x)+b_i(t;\ x)\geqslant 0$.

Для того чтобы в точке x последнее неравенство при условии выполнения (8) выполнялось при всех $i\in\overline{1,\ k}$, $i\equiv I(t;\ x)$, необходимо и достаточно на основании теоремы 1 [5] (для наглядности см. рисунок [5, с. 145]), чтобы для всех $i\in\overline{1,\ k}$, $i\equiv I(t;\ x)$ из условия $-b_i(t;\ x)>B(t;\ x)$ следовало неравенство $\varphi(-b_i(t;\ x))\leqslant 0$.

Отсюда немедленно вытекает справедливость утверждения теоремы.

ЛИТЕРАТУРА

1. *Перекрест В. Т.* Об одной модели одномерного шкалирования // АиТ. 1980. № 2. С. 173–181.
2. *Хмельницкая А. Б.* Об оптимизационных моделях многомерного инвариантного шкалирования // Математическое моделирование и применение вычислительной техники в социологических исследованиях. М.: Ин-т социологических исследований АН СССР, 1980. С. 55–68.
3. *Хмельницкая А. Б.* Проекционные модели инвариантного шкалирования. II // АиТ. 1983. № 7. С. 105–115.
4. *Перекрест В. Т.* Нелинейный типологический анализ социальноэкономической информации. Л.: Наука, 1983.
5. *Хмельницкая А. Б.* Локализация значений характеристической функции шкалирования // АиТ. 1987. № 2. С. 142–151.
6. *Колмогоров А. Н., Фомин С. В.* Элементы теории функций и функционального анализа. М.: Наука, 1968.

Поступила в редакцию
4.VIII.1987

UPDATING THE NOTION OF A SOLUTION IN THE PROBLEM OF INVARIANT SCALING

KHMEL'NITSKAYA A. B.

The condition of correct choice is provided of the deepest stationary point of the functional structural proximity of objects as the scaling function in replacing the original description by an aggregated one.