

Р.Е. САРКИСЯН

ИНТЕРАКТИВНЫЕ ПРОЦЕДУРЫ ИДЕНТИФИКАЦИИ И ОЦЕНИВАНИЯ. Сообщение I. УСЛОВНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ РЕГРЕССИОННЫХ СХЕМ

Մշակված են ինտերակտիվն օգնականների արդյունավետ և անվտանգ լուծումներ գտնելու խնդիրը պարզի տեսակի և ինքնաբերական ծրագիր:

Разработаны интерактивные алгоритм и машинная программа для решения классической задачи условного оценивания параметров регрессионных схем. Библиогр., 5 назв.

An interactive algorithm and machine programming for solving classical problems of conventional evaluation of regression circuit parameters have been developed
Ref. 5

1. *Введение.* Задачи идентификации и оценивания нашли широкое применение в теории и практике управления, системном анализе, экономике, медицине, военном планировании. Они являются также важным звеном всего жизненного цикла проектирования, создания, эксплуатации и совершенствования систем.

В практических приложениях выбор той или иной схемы идентификации и оценивания существенно зависит от вида и объема имеющейся априорной информации о структурных и функциональных особенностях исследуемых и проектируемых систем, их свойствах и связях, а также о тех физических законах, которые определяют внутренние причинно-следственные связи, существование и функционирование систем как единое целое во взаимодействии с окружающей средой.

Определенная связь существует также между априорной информацией и выбором конкретных вычислительных процедур для обработки результатов наблюдения и построения устойчивых оценок и выводов [1].

В настоящем сообщении рассматривается одна из классических задач нахождения оценок по методу наименьших квадратов применительно к регрессионным схемам, когда на оцениваемые параметры наложены определенные функциональные ограничения. В качестве эффективной вычислительной процедуры предложена интерактивная поисковая процедура адаптивного типа, разработанная в рамках методов производных по направлению [2].

2. *Постановка задачи и процедуры регуляризации.* Пусть по результатам наблюдений необходимо построить оценки параметров статистической модели

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_n x_{in} + \varepsilon_i, \quad (1)$$
$$i = 1, 2, \dots, N,$$

в которой x_{ij} , $i=1, \dots, N$; $j=1, \dots, n$ - фиксированные значения независимых переменных; y_i , $i=1, \dots, N$ - наблюдаемые значения зависимой переменной (отклика); β_j , $j=0, 1, \dots, n$ - оцениваемые независимые параметры, значения которых удовлетворяют совместной системе $A\beta = c$, где A - матрица размерности $m \times (n+1)$, c - заданный $(m \times 1)$ вектор, ε_i , $i=1, \dots, N$ - некоррелированные друг с другом случайные величины с нулевым математическим ожиданием и дисперсией σ^2 .

В терминах стандартной задачи математического программирования нахождение условных МНК-оценок неизвестных параметров имеет вид

$$E(\beta) = \varepsilon^T \varepsilon \rightarrow \min_{\beta \in E^*}, \quad A\beta = c, \quad (2)$$

где $\varepsilon = y - X\beta$, $y = (y_1, \dots, y_N)^T$, $\beta = (\beta_0, \beta_1, \dots, \beta_n)^T$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)^T$.

$X = [x_{ij}]$, $i=1, \dots, N$, $j=0, 1, \dots, n$, $x_{i0} = 1, \forall i$.

Как известно [3, 4], с помощью метода Лагранжа можно получить условную оценку в виде $\hat{\beta} = (X^T X)^{-1} (\Lambda^T (X^T X)^{-1} \Lambda)^{-1} (c - Ab) + b$, где $b = (X^T X)^{-1} X^T y$ - безусловная МНК-оценка вектора β . Если матрица $X^T X$ плохо обусловлена, то полная дисперсия оценок $\sum \text{var}(\hat{\beta}_j) = \sigma^2 \text{Tr}(X^T X)^{-1}$ оказывается слишком большой величиной. Обусловленная ею величина $\sum \text{var}(\hat{\beta}_j)$ также будет большой, поэтому условная оценка $\hat{\beta}$ непригодна для практических целей.

Традиционно для уменьшения дисперсии оценок параметров применяются различные процедуры регуляризации, например, типа $(X^T X + kI)^{-1}$, где k - корректирующий параметр, I - единичная матрица, которые приводят к так называемым гребневым оценкам [3]. Последние хотя и не являются несмещенными, однако более устойчивы, чем некорректированные.

Избежать нежелательных эффектов, обусловленных обращением плохо обусловленных матриц, можно и без корректирующих процедур, а путем аппроксимации исходной задачи условного оценивания соответствующей многокритериальной задачей и применения интерактивных поисковых процедур, использующих лишь информацию о градиентах целевых функций и их маргинальных отношениях.

3. Эквивалентная многокритериальная задача. С формально-математической точки зрения условие $A\beta = c$ эквивалентно решению задачи

$$(c - A\beta)^T (c - A\beta) \rightarrow \min_{\beta \in E^*}, \quad (3)$$

поэтому задачу условной оптимизации (1) можно аппроксимировать двухкритериальной задачей математического программирования

$$f(\beta) = (f_1(\beta), f_2(\beta))^T \rightarrow \min_{\beta \in E^*}, \quad (4)$$

где $f_1(\beta) = \varepsilon^T \varepsilon$, $f_2(\beta) = (c - A\beta)^T (c - A\beta)$. Применим к решению задачи общую стратегию адаптивных интерактивных поисковых процедур, основанных на методах производных по направлению [2].

Пусть β^{*1} и β^{*2} - соответствующие решения локальных оптимизационных задач, так что $\beta^{*1} = b = (X^T X)^{-1} X^T y$ и $A\beta^{*2} = c$. Через β^k , $k = 0, 1, \dots$, обозначим некоторое текущее решение и предположим, что один из критериев, например $f_2(\beta)$, выбран в качестве опорного (наиболее важного) критерия в этом состоянии. Его отметим индексом $j_k = 2$, так что $f_{j_k} = f_2$. В состоянии $\beta^k \in E$ градиенты критериальных функций имеют вид

$$S^1 = \nabla f_1(\beta^k) = -2X^T y + 2X^T X \beta^k,$$

$$S^2 = \nabla f_2(\beta^k) = 2A^T c + 2A^T A \beta^k.$$

Если по мнению лица, принимающего решения, приращения Δf_2 , Δf_{j_1} и Δf_{j_2} , $\Delta f_{j_1} < \Delta f_{j_2}$, таковы, что имеет место соотношение $(f_1^k + \Delta f_{j_1}, f_2^k) < (f_1^k, f_2^k + \Delta f_{j_2}) < (f_1^k + \Delta f_{j_2}, f_2^k)$, то маргинальное отношение замещения между критериями f_1 и f_2 определяются в виде $\mu_{12} = S^{11}e / S^{21}e$ и принадлежит интервалу $[A_1, B_1]$, где $A_1 = \Delta f_2 / \Delta f_{j_1}$, $B_1 = \Delta f_2 / \Delta f_{j_2}$, e - направление наискорейшего возрастания функции полезности (точнее, ее условный градиент). Очевидно, что конус предпочтительных направлений $P_0 = \{e \in E^{n+1} / S^{11}e / S^{21}e \in [A_1, B_1]\}$ является непустым подмножеством множества направлений спуска обеих критериальных функций $F_0 = \{d \in E^{n+1} / S^{11}d, S^{21}d \leq 0\}$. К тому же в точке β^k любое направление возможно, т.е. конус аппроксимации G_0 [5] совпадает со всем пространством E^{n+1} . Следовательно, $F_0 \cap G_0 \cap P_0 = P_0$.

Исходя из этих соображений, наилучшее направление поиска находится путем решения задачи

$$\Phi(d) = \{-(S^{11}d)^2 / S^{21}d - S^{2T}d\} \rightarrow \max_{d \in E^{n+1}, \|d\|=1} \quad (5)$$

Обозначим ее решение через $e^k = d^k$, а соответствующую величину шага по e^k , как решение задачи $\arg \min f_2(\beta^k + \sigma e^k)$, $\sigma \geq 0$, через σ_k . Легко установить, что $\sigma_k = (c^T A e^k - e^{kT} A^T A \beta^k) / \|A e^k\|^2$. При этом новое решение будет равно $\beta^{k+1} = \beta^k + \sigma_k \cdot e^k$. Оно, очевидно, удовлетворяет условию $A\beta = c$.

Таким образом, за одну итерацию из произвольной точки β^k попадаем в точку минимума критериальной функции $f_2(\beta)$, что является одним из решений задачи $A\beta = c$.

Во второй итерации в качестве опорного (ведущего) критерия естественно выбрать $f_1(\beta)$, т.е. $j_k = 1$, а в качестве направления движения - направление e^k , все точки $\beta(\sigma) = \beta^k + \sigma e^k$ которого

удовлетворяют условию $A\beta = c$, т.е. имеет место равенство $A(\beta^i + \sigma e^i) = c$. Так как $A\beta = c$, получаем, что $Ae^i = 0$. В терминах целевой функции $f_1(\beta)$ последнее условие эквивалентно условию $S^{-1}e^i = 0$, следовательно, $\mu_{21} = S^{-1}e^i / S^{-1}e^i = 0$, $P_0 = \{e \in E^{n+1} / S^{-1}e / S^{-1}e = 0\}$, $E_0 = \{d \in E^{n+1} / S^{-1}d \leq 0, S^{-1}d = 0\}$, и вектор e^k совпадает с решением задачи

$$\Phi(d) = \{-S^{-1}d\} \rightarrow \max_{d \in E_0, \|d\|=1} \quad (6)$$

Ввиду очевидных свойств задачи (6) она имеет единственное решение.

Оптимальная величина шага данной итерации равна $\sigma_i = \arg \min f_1(\beta^i + \sigma e^i) = (X^i Y e^i - e^i X^i X \beta^i) / \|X e^i\|$, а новое решение принимает вид $\beta^{i+1} = \beta^i + \sigma_i e^i$. Простой проверкой можно убедиться в том, что в точке β^{i+1} имеет место условие $S^{-1}e^i = 0$, т.е. она является точкой условного минимума функции $f_1(\beta)$.

Таким образом, применение итерационной процедуры порождает целенаправленный вычислительный процесс, в результате которого исходная условная задача оценивания решается за две итерации.

Соответствующий вычислительный алгоритм имеет вид:

Шаг 1. Сформулировать задачу (4) и найти ее локальные решения β^1 и β^2 . Если $A\beta^1 = c$, остановиться. В противном случае перейти к шагу 2.

Шаг 2. Положить $k=1$, выбрать $f_1(\beta)$ в качестве опорного критерия и подходящее начальное приближение $\beta^k \in E^{n+1}$. Если $A\beta^k = c$, перейти к шагу 5. В противном случае - к шагу 3.

Шаг 3. Сформировать интервал $[A_k, B_k]$ и решить задачу (5). Пусть e^1 - ее решение, а $\sigma_1 = \arg \min f_1(\beta^k + \sigma e^1)$ при $\sigma \geq 0$.

Шаг 4. Вычислить новое решение $\beta^{k+1} = \beta^k + \sigma_1 e^1$ и положить $k = k + 1$.

Шаг 5. Выбрать f_1 в качестве опорного критерия, сформировать множество F_0 и решить задачу (6). Пусть e^1 - ее решение, а $\sigma^1 = \arg \min f_1(\beta^k + \sigma e^1)$ при $\sigma \geq 0$.

Шаг 6. Вычислить $\beta^{k+1} = \beta^k + \sigma_1 e^1$ и остановиться. Вектор $\beta = \beta^{k+1}$ представляет собой условную МНК-оценку параметров модели (1), при котором $f_1 = f_1(\beta) = \min f_1(\beta)$ при $\beta \in E^{n+1}$, $A\beta = c$.

4. Модельный пример и обсуждение результатов. Рассмотрим задачу оценивания параметров статистической модели $y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i$, $i = 1, \dots, N$, при условии связи $\beta_0 + \beta_1 = 1$. Решим эту задачу при следующих исходных данных [3]:

$$y^T y = 2286.31; X^T y = (\sum y_i, \sum x_i y_i)^T = (235.60, 11821.4320)^T;$$

$$X^T X = \begin{bmatrix} N & \sum x_{ii} \\ \sum x_{ii} & \sum x_{ii}^2 \end{bmatrix} = \begin{bmatrix} 25 & 1315 \\ 1315 & 76323.42 \end{bmatrix}.$$

Ее безусловное решение равно $b = \beta^{**} = (13.62, -0.08)^T$. Пусть $\beta^1 = \beta^k = (0, 0)^T$. Так как в рассмотренном случае $A = [-1, 1]$ и $c = 1$, то для градиентов S^1 и S^2 получаем выражение $S^1 = \nabla f_1(\beta^k) = 2(235.6, 1183.432)^T$; $S^2 = \nabla f_2(\beta^k) = (2, -2)^T$. Для произвольных $0 \leq \lambda_1 < \lambda_2 < \infty$ решение первой итерации равно $e^k = (1/\sqrt{5}, 2/\sqrt{5})^T$, $\sigma_k = \sqrt{5}$, $\beta^k = (1, 2)^T$.

В новом состоянии $S^1 = (4838.4, 284281)^T$, $S^2 = (0, 0)^T$. Выбирая для второй итерации в качестве опорного критерий $f_1(\beta)$ и решая соответствующую подзадачу, получаем $e^1 = (-\sqrt{2}/2, -\sqrt{2}/2)^T$, $\sigma_k = \arg \min f_1(\beta_k + \sigma e^1) = 2.59$, $\beta^{k+1} = (-0.83, 0.17)^T$. Так как $A\beta^{k+1} = c$, а $S^{1T}d = \nabla f_1(\beta^{k+1})^T d \geq 0$ для всех направлений, для которых $Ad = 0$, то найденная точка β^{k+1} является точкой условного минимума функции $f_1(\beta) = \varepsilon^1 \varepsilon$ при условии $A\beta = c$. При этом $f_1(\beta^*) = 509.996$, тогда как $f_1(b) = f_1^* = 20.4291$.

Приведенная вычислительная процедура имеет простую геометрическую интерпретацию: в результате первой итерации из произвольной начальной точки β^1 мы попадаем на поверхность, описываемую условием $A\beta = c$. Движение вдоль нее во второй итерации приводит к искомой точке $\beta^* = \arg \min f_1(\beta)$ при $\beta \in E^{**}$ и $A\beta = c$.

Заключение. Задачи оценивания, в которых априорная информация задана в виде функциональных ограничений, можно интерпретировать в терминах многокритериальных задач математического программирования с последующим применением для их решения интерактивных поисковых процедур адаптивного типа. В качестве дополнительной информации при этом выступает интервальное оценивание маргинальных соотношений между критериями.

ЛИТЕРАТУРА

1. Современные методы систем / Под ред. Э.П. Эйхоффа. - М.: Мир, 1983. - 400 с.
2. Саркисян Р.Е. Адаптивные человеко-машинные процедуры для диалоговых систем. Сообщение 2. Методы производных по направлению // Ин. АН Арм. ССР. Сер. ТН. - Т. 44, № 5-6, - 1991. - С. 282-287.
3. Себер Дж. Линейный регрессионный анализ. - М.: Мир, 1980. - 456 с.
4. Рао С.Р. Линейные статистические методы и их применение. - М.: Наука, 1968. - 347 с.
5. Базар М., Шетти К. Нелинейное программирование. Теория и алгоритмы. - М.: Мир, 1982. - 583 с.