1.Денисов П.А. О проектировании водохранилищ-охладителей // Электрические станции. – 1962. – №2. – С.20-26.

Получено 25.09.2002

УДК 628.94008.6 : 518.9

О.А.ПРОСКУРНИН
УкрНИИЭП, г.Харьков

СРАВНЕНИЕ РЕЗУЛЬТАТОВ РЕШЕНИЯ РЕГРЕССИОННОЙ ЗАДАЧИ МЕТОДОМ МОНТЕ-КАРЛО ПРИ РАЗЛИЧНЫХ БАЗИСАХ ФУНКЦИОНАЛЬНОГО ПРОСТРАНСТВА

Исследуется зависимость результата восстановления функции регрессии при решении экологических задач от выбора базиса функционального пространства. Вводится в рассмотрение критерий качества решения, учитывающий точность и надежность результата. Проводится сравнение результатов решения по данному критерию.

Одним из основных статистических методов обработки натурных данных при решении экологических задач является регрессионный анализ – восстановление функциональной зависимости математического ожидания одной случайной величины от одной или нескольких других случайных величин.

В [1] описывается непараметрический (независимый от вида распределения случайных величин) способ решения задачи регрессии путем применения метода статистических испытаний (метод Монте-Карло). Суть метода заключается в следующем: по данным выборочных наблюдений $\{x_i, y^*(x_i)+\varepsilon_i\}$, $i=1,\ldots,n$, где $y^*(x)$ – истинная функция регрессии, а ε_i – наложенный шум, искомая функция $y(x)$ (являющаяся оценкой $y^*(x)$) представляется как элемент евклидового функционального пространства Φ с заданным скалярным произведением элементов (f_1,f_2) и ортогональным базисом $\{\Theta_k\}$ [5]:

$$y(x) = \sum_{k=0}^{M} \gamma_k \Theta_k , \quad (1)$$

где γ_k – коэффициенты разложения, рассматриваемые в данной задаче как неизвестные параметры регрессии; M – порядок искомой функции, определяемый вместе с γ_k.

Каждый параметр регрессии γ_k исходя из ортогональности базиса равен:

$$\gamma_k = \frac{(y, \Theta_k)}{\|\Theta_k\|^2} \tag{2}$$

и может оцениваться как интегральный показатель по выборке $\{(y_i, \Theta_k)\}$ методом Монте-Карло с последующей проверкой нулевой гипотезы [1]. Множество статистически значимых параметров γ_k определяет, во-первых, саму функцию $y(x)$, во-вторых, границы доверительной полосы:

$$y^-(x) = y(x) - 2\sigma_y, \qquad y^+(x) = y(x) + 2\sigma_y. \tag{3}$$

Как вариант, может применяться механизм последовательного исключения слагаемых [1], когда из ряда наблюдений $\{y_n\}$ перед очередной оценкой вычитаются величины $\sum_k (\gamma_k, \Theta_k)$, соответствующие ранее оцененным слагаемым регрессионной зависимости. В [1, 2] приводятся примеры расчета регрессии вышеописанным методом на натурных данных.

Остается открытым вопрос выбора базиса $\{\Theta_k\}$, который при малых объемах выборки может существенно влиять на результат. В [3] в качестве первого шага в решении данного вопроса производилось ранжирование результатов расчета, выполненных на модельном примере при различных базисах, по критерию качества:

$$C = \frac{1}{2}\frac{S_{4\sigma}}{S_{4\sigma\max}} + \frac{1}{2}\frac{\mu(Dош)}{\mu(D)}, \tag{4}$$

где μ – функционал меры; D – область определения функции $y(x)$; $D_{ош} \subset D$ – множество точек, в которых выполняется условие $(y^*(x) > y(x) + 2\sigma) \vee (y^*(x) < y(x) - 2\sigma)$; $S_{4\sigma}$ – площадь доверительной полосы (3), равная $4\int_D \sigma_y dx$; $S_{4\sigma max}$ – максимальная площадь доверительной полосы в рассматриваемом ряду экспериментов.

Недостатком данного критерия является зависимость величины C от $S_{4\sigma max}$, что приводит к зависимости результата ранжирования от результата расчета с наибольшим $S_{4\sigma}$. Кроме того результат ранжирования зависел от случайной выборки $\{\varepsilon_i\}$.

Нами предлагается в качестве критерия качества рассматривать интегральную удаленность истинной функции регрессии $y^*(x)$ от границ доверительной полосы $y(x)\pm 2\sigma_y$:

$$I = \int\limits_D [|y^+(x) - y^*(x)| + |y^-(x) - y^*(x)|]dx. \qquad (5)$$

Величина I учитывает как точность результата расчета, так и его надежность, поскольку как уменьшение точности (расширение доверительной полосы), так и уменьшение надежности (увеличение степени выхода за доверительные границы) приводит к возрастанию величины I.

С целью уменьшения случайного фактора, для каждого варианта расчета, обусловленного объемом выборки n, базисом $\{\Theta_k\}$, и характеристикой шума ε, случайным образом (по закону равномерного распределения) задается большое число раз выборка независимых переменных $\{x_i\}$, а также выборка значений шума $\{\varepsilon_i\}$ в узловых точках из конкретной генеральной совокупности. Для каждой пары выборок $\{x_i\}$ и $\{\varepsilon_i\}$ производится решение регрессионной задачи различными способами с последующим расчетом величины I. Сравнение результатов производится по средней величине:

$$\hat{I} = \frac{1}{N} \sum_{\{x_i\},\{\varepsilon_i\}} I(\{x_i\},\{\varepsilon_i\}), \qquad (6)$$

где N – число испытаний.

В настоящей работе приводится сравнение результатов восстановления модельной функции регрессии $y^*(x)=1+x+x^2+x^3$, $x\in[-1,1]$ методом Монте-Карло (с применением механизма последовательного исключения слагаемых) и методом наименьших квадратов (МНК) [3]. Выбор данной модельной функции объясняется равным значением коэффициентов при степенях x, а 3-й порядок выбран исходя из того, что в реальных экологических задачах с малыми объемами выборки и значительным шумом получение статистически значимого слагаемого 4-го порядка маловероятно. В качестве базиса $\{\Theta_k\}$ использовались различные ортогональные системы полиномов (таблица).

Результат расчета величины I (средней по 100 испытаниям) для объемов выборки $n\in\{10;13;16\}$ и равномерном шуме интенсивности $\sigma_\varepsilon=\{0,5;0;9\}$ в виде диаграмм приведен на рис.1, 2 (меньшей величине I соответствует лучший результат решения). Как видно из диаграмм, во всех рассматриваемых случаях метод Монте-Карло оказывается более эффективным при использовании в качестве базиса системы полиномов Чебышева 2-го рода. Кроме того, уменьшение объема выборки и увеличение интенсивности шума снижают преимущество МНК по сравнению с методом Монте-Карло по рассматриваемому критерию. (Для шума σ_ε=0,5 МНК превосходит метод Монте-Карло при n=13 и n=16, а для шума σ_ε=0,9 незначительно превосходит лишь при n=16). Таким образом, проведенный эксперимент демонстрирует эффективность применения регрессионного метода Монте-Карло на базе системы полиномов Чебышева 2-го рода при малом объеме выборки и значительном шуме.

Ортогональные системы полиномов

Название системы полиномов	**Общепринятое обозначение и формула k-го члена,**	**Область ортогональности**	**Скалярное произведение, обеспечивающее ортогональность**	**Квадрат нормы**
Лежандра	$P_k = \frac{1}{2^n(n)!}\frac{d^n}{dx^n}(x^2-1)$	[-1,1]	$(P_kP_l) = \int_{-1}^{1} P_kP_l dx$	$\frac{2}{2k+1}$
Чебышева 1-рода	$T_k(x) = \cos(k\arccos(x))$	[-1,1]	$(T_kT_l) = \int_{-1}^{1}\frac{T_kT_l}{\sqrt{1-x^2}}dx$	$\frac{\pi}{2}$, $k\neq0$, π, k=0
Чебышева 2-рода	$U_k(x) = \frac{\sin[(k+1)\arccos(x)]}{\sqrt{1-x^2}}$	[-1,1]	$(U_kU_l) = \int_{-1}^{1} U_kU_l\sqrt{1-x^2}dx$	$\frac{\pi}{2}$
Ортонормируемые по сумме	$L_k(x)^*$	$\{x_i\}$	$(L_kL_l) = \sum_{i=1}^{n} L_k(x_i)L_l(x_i)$	1

* – ортонормируемые по сумме полиномы не выражаются единой формулой; алгоритм их вычисления подробно описан в [3].

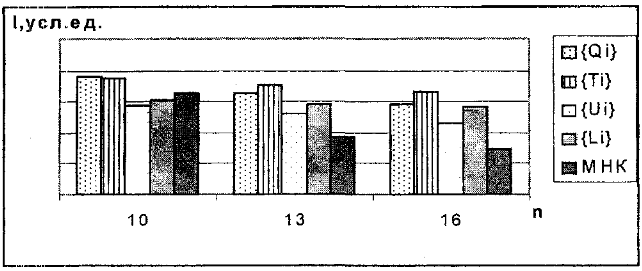

Рис.1 – Сравнение результатов решения регрессионной задачи при σ_ε=0,5

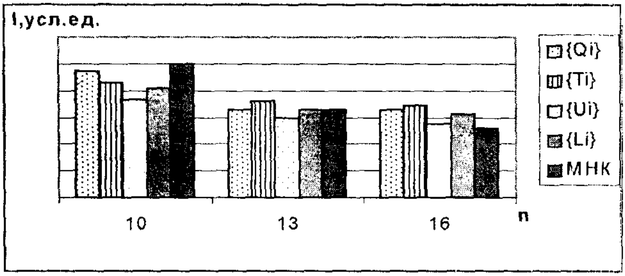

Рис.2 – Сравнение результатов решения регрессионной задачи при σ_ε=0,9

1.Баранник В.А., Проскурнин О.А. Применение метода статистических испытаний в регрессионном анализе данных экологических исследований // Экологическая, техногенная безопасность и социальный прогресс: Вестник ХИСП. Вып.1. – 2001.

2.Баранник В.А. Непараметрический метод оценки точности и достоверности определения интегральных показателей по данным выборочных наблюдений // Проблемы охраны окружающей природной среды и техногенной безопасности: Сб. научн. тр. – Харьков: УкрНИИЭП, 2000. – С.129-136.

3.Проскурнин О.А. Численный подход к оценке качества решения регрессионной задачи при различных базисах функционального пространства // Экологическая, техногенная безопасность и социальный прогресс: Вестник ХИСП. Вып.1 (2). – 2002.

4.Худсон Д. Статистика для физиков. – М.: Мир, 1970.

5.Ильин В.А. Линейная алгебра. – М.: Наука, 1974.

Получено 27.09.2002

АРХИТЕКТУРА

УДК 72.03

И.В.ДРЕВАЛЬ
Харьковская государственная академия городского хозяйства

ЭКОЛОГИЧЕСКИЕ АСПЕКТЫ АРХИТЕКТУРНО-ГРАДОСТРОИТЕЛЬНОГО ФОРМИРОВАНИЯ ЭЛЕМЕНТОВ ТРАНСПОРТНО-КОММУНИКАЦИОННОГО КАРКАСА УРБАНИЗИРОВАННОЙ СРЕДЫ

Рассматривается влияние экологических ценностей на структурное и композиционное формообразование элементов транспортного каркаса урбанизированной среды.

Целенаправленное формирование эстетически и экологически полноценной окружающей среды – одна из важнейших задач совре-