УДК 519.8

ОБ ОДНОЙ ЗАДАЧЕ ИДЕНТИФИКАЦИИ СЛОЖНЫХ СИСТЕМ

Дмитриев Е. И.

научный руководитель проф. Медведев А. В.

Сибирский государственный аэрокосмический университет имени академика М.Ф. Решетнева

В задаче идентификации сложных систем в условиях непараметрической неопределенности исследователь располагает некоторой исходной обучающей выборкой. Часто в пространстве «входных-выходных» переменных элементы этой выборки образуют «сгущения», а в ряде других подобластей пространства-«разряжения». При их использовании при решении задачи идентификации предлагается из имеющейся обучающей выборки сгенерировать новую «рабочую» выборку, которая непосредственно используется в непараметрических моделях. Этот процесс генерации, предлагаемый в настоящем докладе целесообразно использовать и в других непараметрических адаптивных системах. Процесс генерации основывается на непараметрическом оценивании регрессионных характеристик по наблюдениям с шумами.

Если осуществить процесс генерации новой «рабочей» выборки из исходной обучающей $(x_i, u_i, i = \overline{1,s})$ на основе создания равномерной сетки в пространстве $\Omega(u)$и вычисления в ее узлах непараметрической оценки$x \epsilon\ \Omega(x)$, то можно получить другую «рабочую» выборку $(x_i', u_i', i = \overline{1,s'})$, где $s' \ll s$. На этой выборке вводится новый тип стохастических аппроксимаций из класса Н-аппроксимаций.Вычислительные эксперименты показывают, что объем рабочей выборки, при практически том же качестве оценивания регрессионных характеристик, значительно меньше исходной обучающей выборки. Для генерации элементов «рабочей» обучающей выборки используется соответствующий непараметрический индикатор.

Целесообразно рассмотреть алгоритм генерации новой «рабочей» выборки более подробно:

1. После этапа сбора информации, снятия некоторых измерений, связанных с работой некоторой сложной системой, получения исходной выборки, производится общий анализ данной выборки (установление размерности, выявление «сгущений» и «разряжений», наличия шумов и т.д.).

2. Далее происходит настройка значений параметров размытости c_s^j с помощью метода скользящего экзамена, где s- объем исходной выборки, $j = \overline{1,p}$, p- размерность пространства измерений.

3. Затем происходит построение непараметрической оценки Надарая-Ватсона с использованием ранее настроенных значений параметров размытости:

$$\hat{x}_s(u) = \frac{\sum_{i=1}^{s} x_i \cdot \prod_{j=1}^{p} \Phi\left(\frac{u^j - u_i^j}{c_s^j}\right)}{\sum_{i=1}^{s} \prod_{j=1}^{p} \Phi\left(\frac{u^j - u_i^j}{c_s^j}\right)},$$

где $\Phi(\cdot)$- колоколообразное ядро, x_i- значения выхода, u_i^j- значения входа, $j = \overline{1,p}$, $i = \overline{1,s}$, p- размерность, s- объем выборки.

4. Затем происходит построение некоторой равномерной сетки объемом s''.

5. Происходит «отсеивание» некоторых узлов данной равномерной сетки с помощью индикатора, который является знаменателем непараметрической оценки, формула которой приведена выше. То есть индикатор имеет следующий вид:

$$I(u^j) = \sum_{i=1}^{s} \prod_{j=1}^{p} \Phi\left(\frac{u^j - u_i^j}{c_s^j}\right).$$

Индикатором отсеиваются те точки равномерной сетки, значение индикатора в которых равно 0.

6. Таким образом, после работы индикатора получается выборка входов объемом s'. В оставшихся узлах сетки вычисляются значения выхода с помощью уже упомянутой ранее непараметрической оценки Надарая-Ватсона. В результате алгоритма сгенерирована новая «рабочая» выборка.

7. Далее целесообразно использовать аппроксимацию по данной выборке с помощью Н-аппроксимации:

$$\hat{x}_{s'}^{H}(u) = \frac{1}{s' \cdot \prod_{j=1}^{p} c_{s'}^j} \sum_{i=1}^{s'} x_i' \cdot \prod_{j=1}^{p} H^j\left(\frac{u^{j'} - u_i^{j'}}{c_{s'}^j}\right),$$

$$H^j(z) = \left(b^j - a^j\right) \cdot \Phi(z),$$

где $c_{s'}^j$ - новые значения параметров размытости, x_i'- значения выходов в новой «рабочей» выборке, $H(\cdot)$- Н-ядро (содержит внутри себя ядро и длины областей определения для каждой переменной), $u_i^{j'}$ - значения входов в новой «рабочей» выборке.

Ниже приведен пример генерации рабочей выборки для размерности $p = 2$.

Истинная зависимость $f(x,y) = 0.2x + 0.1y$. Общий объем выборки (включая «сгущения») из области определения $x \in [0;20]$ и $y \in [0;20]$ $s = 2600$. Количество «сгустков» $N_{sg} = 3$ объемом $s_{sg} = 200$. На выборку наложена 10%-помеха, распределенная по равномерному закону.

Выборка на плоскости $0xy$ выглядит следующим образом (рис. 1):

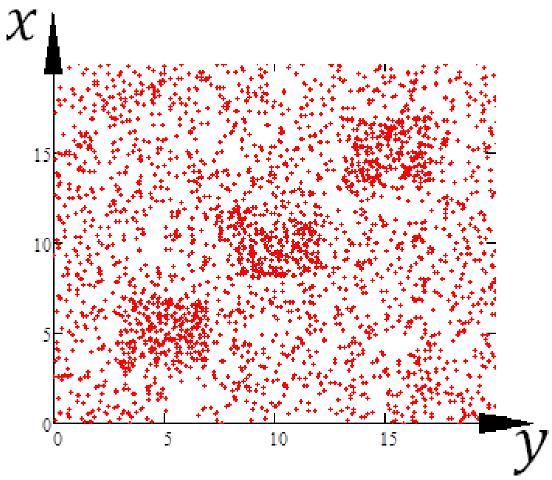

Рис. 1. Выборка.

В результате работы алгоритма и индикатора объем новой «рабочей» выборки составил $s' = 2169$. Далее в оставшихся точках сетки были вычислены значения выхода, была построена Н- аппроксимация. И при этом ошибка аппроксимации имела тот же порядок малости, что и при построении оценке на исходной выборке, несмотря на значительно сократившийся объем выборки. Ниже приведены графики, на которых сравнивается исходная выборка со сгенерированной «рабочей» (рис. 2а, 2б):

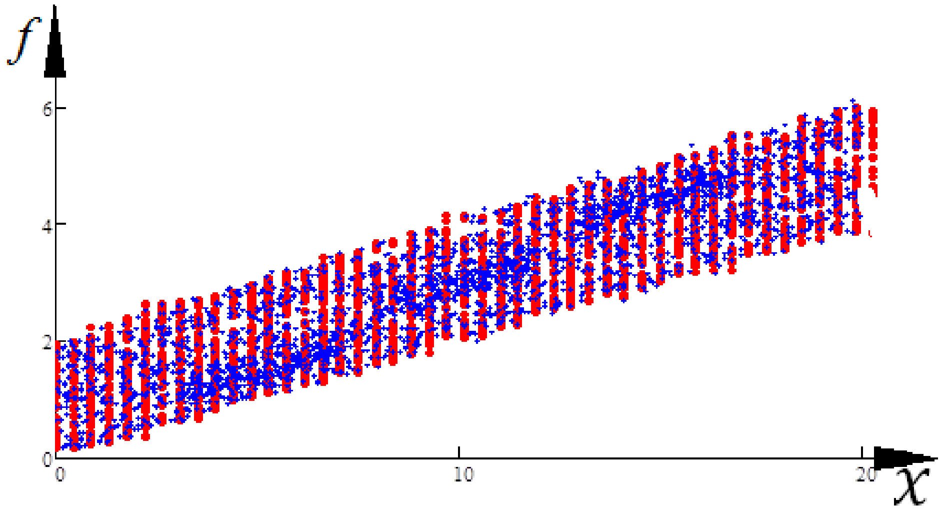

Рис. 2а. Плоскость $f0x$.

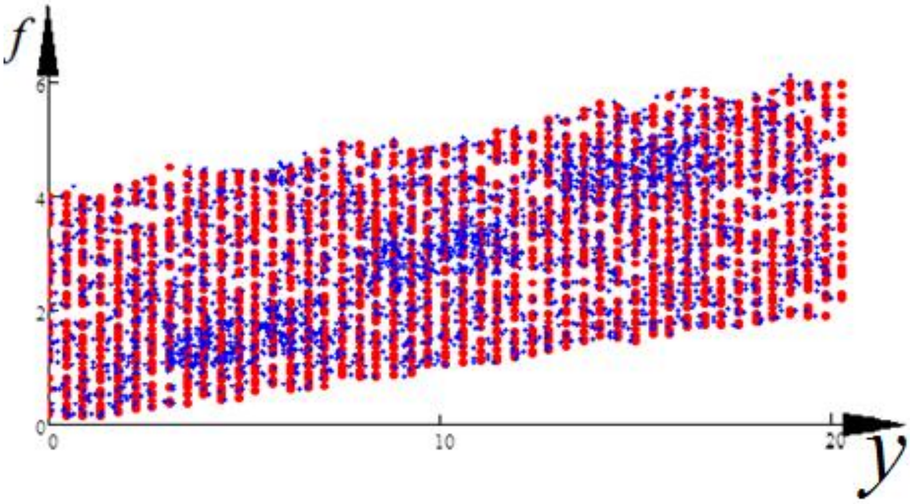

Рис. 2а. Плоскость $f0y$. Исходная выборка представлена красным цветом, сгенерированная- синим.

Таким образом, при значительном уменьшении объема выборки (варьируется кол-во узлов сетки и происходит отсеивание индикатором некоторых точек сетки, в которых знаменатель классической оценки равен 0), не наблюдается значительного увеличения ошибки идентификации (ошибка сохраняет свой порядок малости). Таким образом, в дальнейшей работе для исследований лучше брать новую «рабочую» выборку, т.к. исследователю выгоднее использовать выборку меньшего объема для уменьшения вычислительных затрат.

В статье рассматривается так же случай оценивания регрессионных характеристик при неравномерно распределенных элементах исходной обучающей выборки. Для этого была введена и исследована непараметрическая оценка кривой регрессии по наблюдениям с шумами, использующая в вычислительной формуле вместо измерения выхода объекта x_i $(i = \overline{1,s})$ гиперплоскость $\varphi(x, \overrightarrow{x_k}, \overrightarrow{u_k})$($\overrightarrow{x_k}, \overrightarrow{u_k}$- временные векторы объемом k), аппроксимирующую некоторую окрестность точки (x_i, u_i) (окрестность состоит из точек, выбираемых из соображения их близости к(x_i, u_i) относительно некоторого расстояния h, определяемого экспериментальным путем):

$$\hat{x}_s^{\varphi}(u) = \frac{\sum_{i=1}^{s} \varphi(x, x_i, \overrightarrow{u_i}) \cdot \prod_{j=1}^{p} \Phi\left(\frac{u^j - u_i^j}{c_s^j}\right)}{\sum_{i=1}^{s} \prod_{j=1}^{p} \Phi\left(\frac{u^j - u_i^j}{c_s^j}\right)}$$

Таким образом, каждой точке выборки $(x_i, \overrightarrow{u_i})$ поставлена в соответствие некая гиперплоскость, параметры которой определяются исходя из МНК. В результате исследований данной оценки ошибка идентификации по сравнению с классической непараметрической оценкой имеет значение на порядок, а в некоторых случаях и на два порядка, меньше. Данный результат объясняется привлечением дополнительной полезной информации о поведении соседних с $(x_i, \overrightarrow{u_i})$ точек.

Ниже приведен пример такой модифицированной непараметрической оценки для размерности $p = 1$.

Истинная зависимость: $f(x) = \sin(0.15x)$. Выборка входа была получена на области определения $x \in [0; 10]$ случайным образом (генератор случайных чисел, равномерный закон). Объем выборки составил $s = 30$.

Выборка на плоскости выглядит следующим образом (рис. 3):

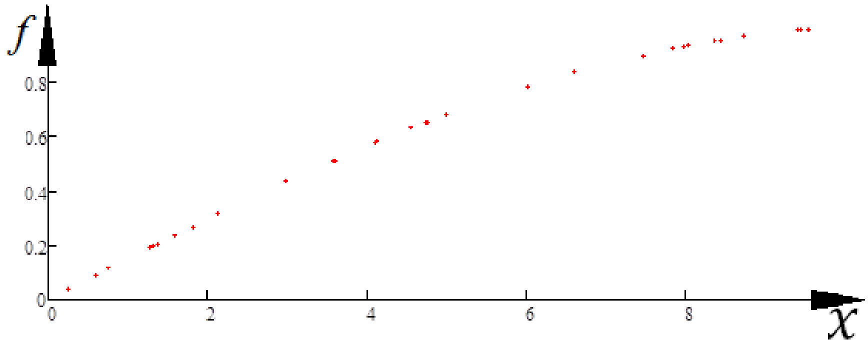

Рис. 3. Выборка.

В результате работы данной модифицированной оценки была получена ошибка: $\varepsilon = 6.699 \cdot 10^{-6}$. Ниже приведен график этой оценки в сравнении с выборкой (рис. 4а):

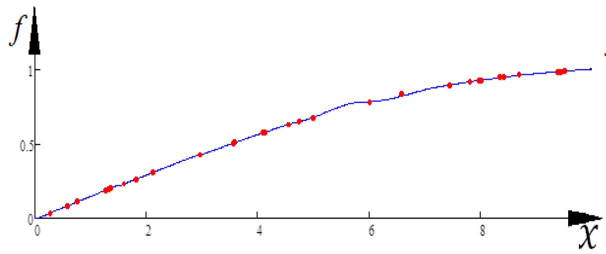

Рис. 4а.

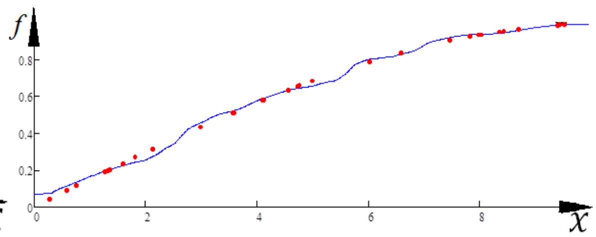

Рис. 4б.

В результате работы обычной непараметрической оценки была получена ошибка: $\varepsilon = 2.362 \cdot 10^{-4}$, что на несколько (2 порядка) больше чем у модифицированной оценки. Выше приведен график этой оценки в сравнении с выборкой (рис. 4б):

Таким образом, данная модифицированная оценка, использующая МНК, дает неплохие результаты даже при работе с малым объемом выборки.