СТАТИСТИЧЕСКАЯ МОДЕЛЬ ПРОСТРАНСТВЕННО РАСПРЕДЕЛЁННОЙ ВРЕМЕННОЙ СИСТЕМЫ

Лапко А.В., Лапко В.А.
Институт вычислительного моделирования СО РАН, Россия, lapko@icm.krasn.ru

Исследуется пространственно распределённый временной процесс

$$y(x(t),\, z) = \psi_z\,(y(t-1),\, x(t),\, z)\,, \tag{1}$$

динамика выходной переменной $y(x(t),\, z) \in R^1$ которого в точке с координатами $z \in R^2$ в момент времени t определяется её предыдущим значением $y(t-1)$ и значениями факторов $x(t) \in R^k$.

Вид преобразования $\psi_z(\cdot)$, зависящего от пространственных координат z, априори неизвестен.

Исходную информацию составляют статистические данные $V_j = \left(x(t),\, y(t),\, t = \overline{1,\, n_j}\right)$, $j = \overline{1,\, n}$ о временной динамике процесса (1) в n точках пространственных координат z^j, $j = \overline{1,\, n}$.

Подобные условия часто встречаются, например, при моделировании экологических, медико-биологических систем и развитии сельскохозяйственных культур.

С позиций принципов коллективного оценивания и методов непараметрической статистики предлагается «обход» проблемы математического моделирования однородных и неоднородных пространственно распределённых временных процессов.

Непараметрические модели однородных процессов. Идея разработанного подхода состоит в построении статистических моделей временных процессов

$$y\left(x(t),\, z^j\right) = \psi\left(y(t-1),\, x(t),\, z^j\right) = \varphi_j\,(y(t-1),\, x(t)),\quad j = \overline{1,\, n}\,, \tag{2}$$

соответствующих пространственным координатам z^j, $j = \overline{1,\, n}$ наблюдения за процессом (1) и последующей их интеграцией в единой решающей функции с помощью непараметрической оценки условного математического ожидания.

Для восстановления временных зависимостей (2) предлагается использовать методику синтеза непараметрических моделей коллективного типа, обеспечивающих наиболее полный учёт информации обучающих выборок V_j на основе сочетания преимуществ параметрических и локальных аппроксимаций [1].

Поставим в соответствие каждому наблюдению $\left(x^t = x(t),\, y^t = y(t)\right) \in V_j$ подвыборку $V_j(\tau) = \left(x^t\,,\, y^t\,,\, t = \overline{\tau+1,\, n_j}\right)$ и линейный полином $\overline{\varphi}_j^{\,\tau}\left(y(t-1),\, x(t),\, \alpha_j^\tau\right)$, параметры которого удовлетворяют условиям

$$y^\tau = \overline{\varphi}_j^{\,\tau}\left(y(\tau-1),\, x(\tau),\, \alpha_j^\tau\right),$$

$$\overline{\alpha}_j^{\,\tau} = \arg\min_{\alpha_j^\tau}\, \left(n_j - \tau\right)^{-1} \sum_{i=\tau+1}^{n_j} \left(y^i - \overline{\varphi}_j^{\,\tau}\left(y^{i-1},\, x^i\,,\, \alpha_j^\tau\right)\right)^2 .$$

Функции $\overline{\varphi}_j^{\,\tau}\left(y(t-1),\, x(t),\, \overline{\alpha}_j^{\,\tau}\right)$ проходят через опорные точки $\left(y^{\tau-1},\, x^\tau,\, y^\tau\right)$ и близки в среднеквадратическом к элементам выборки $V_j(\tau)$, $\tau = \overline{1,\, n_j - k}$.

На основе полученной системы опорных функций $\overline{\varphi}_j^{\tau}\left(y(t-1),\, x(t),\, \overline{\alpha}_j^{\tau}\right)$, $\tau = \overline{1,\, n_j - k}$ построим приближение зависимости

$$y\left(x(t),\, z^j\right) = \varphi_j\left(y(t-1),\, x(t)\right)$$

в виде непараметрической модели коллективного типа

$$\overline{\overline{y}}_j = \overline{\overline{\varphi}}_j\left(y(t-1),\, x(t)\right) = \sum_{i=1}^{n_j-k} \overline{\varphi}_j^i\left(y(t-1),\, x(t),\, \overline{\alpha}_j^i\right)\beta_i\left(y(t-1),\, x(t)\right), \quad j = \overline{1,\, n}, \tag{3}$$

где

$$\beta_i\left(y(t-1),\, x(t)\right) = \frac{\Phi\left(\dfrac{y(t-1)-y^i}{c_\circ}\right)\prod_{v=1}^{k}\Phi\left(\dfrac{x_v(t)-x_v^i}{c_v}\right)}{\sum_{i=1}^{n_j-k}\Phi\left(\dfrac{y(t-1)-y^i}{c_\circ}\right)\prod_{v=1}^{k}\Phi\left(\dfrac{x_v(t)-x_v^i}{c_v}\right)};$$

$\Phi(\cdot)$ - ядерные функции, удовлетворяющие условиям положительности, симметричности и нормированности [2].

Непараметрическая модель коллективного типа (3) характеризуется высоким уровнем помехозащищённости, что обеспечивается двойным сглаживанием при формирования системы упрощённых аппроксимаций $\overline{\varphi}_j^{\tau}(\cdot)$, $\tau = \overline{1,\, n_j - k}$ и путём их усреднения в соответствии с процедурой (3).

Оптимизация непараметрической модели (3) по параметрам размытости $c_\circ = c\, \overline{\sigma}_{t-1}$, $c_v = c\, \overline{\sigma}_v$ ($\overline{\sigma}_{t-1}$, $\overline{\sigma}_v$, $v = \overline{1,\, k}$ - оценки среднеквадратических отклонений $y(t-1)$ и $x_v(t)$, $v = \overline{1,\, k}$) осуществляется из условия минимума эмпирического критерия

$$\overline{\overline{W}}_j(c) = \frac{1}{n_j}\sum_{i1=1}^{n_j}\left(y^{i1} - \overline{\overline{\varphi}}_j\left(y(i1-1),\, x(i1)\right)\right)^2, \tag{4}$$

отражающего меру близости между экспериментальными данными y^{i1} и результатами их оценивания по модели (3).

Выбор параметров c осуществляется в режиме «скользящего экзамена»: ситуация $\left(y^{i1-1},\, x^{i1},\, y^{i1}\right)$, представляемая на контроль, исключается из процесса обучения в процедуре (3), что выполняется при $i \neq i1$.

Для построения обобщённой статистической модели процесса (1) в однородных условиях воспользуемся оператором условного математического ожидания в пространстве z, непараметрическая оценка которого имеет вид

$$\overline{y}\left(x(t),\, z\right) = \overline{\psi}\left(y(t-1),\, x(t),\, z\right) = \frac{\sum_{j=1}^{n}\overline{\overline{\varphi}}_j\left(y(t-1),\, x(t)\right)\beta^j\left(z,\, z^j\right)}{\sum_{j=1}^{n}\beta^j\left(z,\, z^j\right)}, \tag{5}$$

где

$$\beta^j\left(z,\, z^j\right) = \prod_{v=1}^{2}\Phi\left(\frac{z_v - z_v^j}{c_v}\right) \tag{6}$$

- многомерная ядерная функция в пространстве $z = (z_1,\, z_2)$.

С целью повышения аппроксимационных свойств непараметрической модели (5) предлагается ввести в структуру ядерной функции (6) показатели эффективности $\overline{\overline{W}}_j$, $j=\overline{1, n}$ статистик (3), т.е.

$$\beta^j\left(z, z^j\right)=\prod_{v=1}^{2}\Phi\left(\frac{z_v-z_v^j}{c_v}\right)\Phi\left(\frac{0-\overline{\overline{W}}_j}{c}\right), \tag{7}$$

где, например,

$$\Phi\left(\frac{0-\overline{\overline{W}}_j}{c}\right)=\begin{cases}0, \text{ } если \quad \overline{\overline{W}}_j>c \\ \frac{1}{2}, \text{ } если \quad \overline{\overline{W}}_j\le c .\end{cases}$$

Ядерная функция (7) позволяет компенсировать влияние «неудачных» моделей (3) в условиях $\left(y(t-1), x(t), z\right)$, тем самым повышается устойчивость и помехозащищённость процедуры формирования решений (5).

Оптимизация обобщенной модели однородного пространственно распределённого временного процесса (5) по параметрам размытости ядерных функций (6), (7) осуществляется из условия минимума статистического критерия

$$\overline{W}(c)=n^{-1}\sum_{j=1}^{n}n_j^{-1}\sum_{t=1}^{n_j}\left(\overline{\overline{\varphi}}_j\left(y(t-1), x(t)\right)-\overline{\psi}\left(y(t-1), x(t), z^j\right)\right)^2, \tag{8}$$

который определяет оценку среднеквадратической меры близости между частными моделями $\overline{\overline{\varphi}}_j(\cdot)$ при различных условиях $\left(y(t-1), x(t), z^j\right)$, $t=\overline{1, n_j}$, $j=\overline{1, n}$ и их непараметрическим коллективом $\overline{\psi}(\cdot)$.

Непараметрические модели неоднородных процессов. Полученные результаты обобщены на случай неоднородных пространственно распределённых временных систем. Предлагаемая модель имеет двухуровневую структуру.

На первом уровне структуры в соответствии с решающим правилом $m(z)$ определяется принадлежность контрольной ситуации $\left(y(t-1), x(t), z\right)$ к конкретной области однородности Ω_λ, $\lambda=\overline{1, M}$ изучаемого процесса. На втором уровне на основе статистической модели типа (5) оценивается значение выходной переменной $\overline{y}\left(y(t-1), x(t), z\right)$. При построении каждой модели второго уровня используется вся исходная информация $V_j=\left(x(t), y(t), t=\overline{1, n_j}\right)$, $j=\overline{1, n}$. Однако её оптимизация осуществляется по части исходных данных, принадлежащих соответствующей области однородности изучаемого процесса.

Предлагаемый подход развивает результаты исследований, изложенных в работе [3].

Литература

1. Лапко В.А. Непараметрические коллективы решающих правил. - Новосибирск: Наука, 2002 – 168 с.
2. Parzen E. On estimation of a probability density function and mode // Ann. Math. Statistic, 1962, Vol.33. – p. 1065-1076.
3. Лапко А.В., Лапко В.А. Непараметрические системы обработки неоднородной информации. – Новосибирск: Наука, 2007. - 197 с.