РОССИЙСКАЯ АКАДЕМИЯ НАУК

СИБИРСКОЕ ОТДЕЛЕНИЕ

АВТОМЕТРИЯ

№ 1 2002

УДК 519.7

В. А. Лапко

(Красноярск)

НЕПАРАМЕТРИЧЕСКИЕ МОДЕЛИ ВРЕМЕННЫХ ЗАВИСИМОСТЕЙ, ОСНОВАННЫЕ НА МЕТОДЕ ДВОЙНОГО КОЛЛЕКТИВНОГО ОЦЕНИВАНИЯ*

Разработаны непараметрические модели временных зависимостей, обеспечивающие эффективное использование информации коротких рядов наблюдений их параметров с позиций принципов коллективного оценивания. Исследованы асимптотические свойства моделей и проведено их сравнение.

Введение. Объектом исследования является временная зависимость

$$y(t)=\psi_t(x(t)), \qquad (1)$$

где преобразование $\psi_t(\cdot)$ при каждом $t \in T$ имеет, по крайней мере, две первые ограниченные производные по $x_\nu(t)$, $\nu=\overline{1,k}$. Априорная информация о временной зависимости (1) представлена в выборке статистически независимых наблюдений $V=(y^t=y(t),\ x^t=x(t),\ t=\overline{1,n})$.

Для восстановления зависимости (1) воспользуемся методикой синтеза непараметрических моделей коллективного типа, основанной на построении системы ее упрощенных аппроксимаций относительно временной последовательности элементов выборки V, которые объединяются в коллектив решающих правил с помощью методов непараметрической статистики [1]. В предлагаемом подходе системы упрощенных аппроксимаций формируются в двух направлениях от начального $(x^1,\ y^1)$ и конечного $(x^n,\ y^n)$ условий, что позволяет на их основе построить две непараметрические модели коллективного типа. Организация последних в обобщенной модели зависимости (1) осуществляется в виде линейного функционала, тем самым реализуется принцип двойного коллективного оценивания (непараметрический и параметрический). Особенность подобного класса моделей состоит в максимальном использовании исходной статистической информации, что необходимо при анализе коротких временных рядов.

* Работа выполнена при поддержке Красноярского краевого фонда науки (грант № 10F0023M) и Российского гуманитарного научного фонда (грант № 98-06-12001в).

1. **Синтез непараметрических моделей временных зависимостей.** Методика построения модели, основанная на принципе двойного коллективного оценивания, представляется следующим алгоритмом.

Принять $t = 1$.

1. Сформировать выборку $V_t = (x^\tau,\ y^\tau,\ \tau = \overline{t+1, n}) \subset V$. Построить упрощенную аппроксимацию $\varphi_t(x(t), \alpha_t)$ зависимости (1) с учетом условий

$$y^t = \varphi_t(x^t,\ \alpha_t), \qquad \overline{\alpha}_t = \arg\min_{\alpha_t} \sum_{(x^\tau,\ y^\tau) \in V_t} (y^\tau - \varphi_t(x^\tau,\ \alpha_t))^2. \tag{2}$$

В соответствии с приведенными условиями упрощенная аппроксимация проходит через точку $(x^t,\ y^t)$ и близка в среднеквадратическом ко всем последующим ситуациям $(x^\tau,\ y^\tau) \in V_t$.

Если $t < n - k$, перейти к этапу 1 при $t = t + 1$, в противном случае – к этапу 2.

2. Построить непараметрическую модель коллективного типа зависимости (1):

$$\bar{y}(t) = \overline{\psi}_t(x(t)) = \sum_{\tau=1}^{n-k} \varphi_\tau(x(t), \overline{\alpha}_\tau)\beta_\tau(x(t)), \tag{3}$$

где

$$\beta_\tau(x(t)) = \frac{\prod\limits_{\nu=1}^{k} \Phi\left(\dfrac{x(t) - x^\tau}{c}\right)}{\sum\limits_{\tau=1}^{n-k} \prod\limits_{\nu=1}^{k} \Phi\left(\dfrac{x(t) - x^\tau}{c}\right)},$$

а ядерные функции $\Phi(\cdot) \in H$ удовлетворяют условиям нормированности, положительности и симметричности [1] относительно опорных ситуаций $x^\tau,\ \tau = \overline{1, n-k}$.

Принять $t = n$.

3. Сформировать выборку $V_t' = (x^\tau,\ y^\tau,\ \tau = \overline{1, t-1})$. Построить упрощенную аппроксимацию $\varphi_t'(x(t), \alpha_t')$, проходящую через точку $(x^t,\ y^t)$, параметры которой удовлетворяют дополнительному условию

$$\overline{\alpha}_t' = \arg\min_{\alpha_t'} \sum_{(x^\tau,\ y^\tau) \in V_t'} (y^\tau - \varphi_t'(x^\tau,\ \alpha_t'))^2.$$

Если $t > k$, перейти к этапу 3 при $t = t - 1$, иначе – к этапу 4.

4. Систему упрощенных аппроксимаций $\varphi_t'(x(t), \alpha_\tau'),\ \tau = \overline{k, n}$, организовать в непараметрический коллектив

$$\bar{y}_t(t) = \overline{\psi}_t'(x(t)) = \sum_{\tau=k}^{n} \varphi_\tau'(x(t), \overline{\alpha}_\tau')\beta_\tau(x(t)). \tag{4}$$

Построить обобщенную коллективную модель зависимости (1) в соответствии с процедурой

$$\bar{\bar{y}}(t) = \beta \bar{\psi}'_t(x(t)) + (1-\beta)\bar{\psi}_t(x(t)), \qquad (5)$$

где параметр β определяется из условия минимума эмпирического критерия

$$\bar{\bar{y}}(t) = \bar{\psi}_t(x(t)) + \beta(\bar{\psi}'_t(x(t)) - \bar{\psi}_t(x(t))), \qquad (6)$$

а упрощенные аппроксимации в форме

$$\varphi_t(x(t), \alpha_t) = y^t + (\varphi_t(x(t)) - \varphi_t(x^t, \alpha_t)), \qquad (7)$$

$$\varphi'_t(x(t), \alpha'_t) = y^t + (\varphi'_t(x(t)) - \varphi'_t(x^t, \alpha_t)), \qquad (8)$$

используя условия типа (2) их прохождения через точки (x^t, y^t). Здесь $\varphi'_t(x(t))$ и $\varphi_t(x(t))$ – упрощенные аппроксимации без свободного члена.

Подставим (7), (8) с учетом статистик (3), (4) в модель (6), получим

$$\bar{\bar{y}}(t) = \sum_{\tau=1}^{n-1} y^\tau \beta_\tau(x(t)) - \beta\left(\sum_{\tau=1}^{k-1} y^\tau \beta_\tau(x(t)) - \sum_{\tau=n-k+1}^{n} y^\tau \beta_\tau(x(t))\right) +$$

$$+\beta\left(\left(\frac{1}{\beta}-1\right)\sum_{\tau=1}^{n-k}(\varphi_\tau(x(t)) - \varphi_\tau(x^\tau, \alpha_\tau))\beta_\tau(x(t)) +\right.$$

$$\left. + \sum_{\tau=k}^{n}(\varphi'_\tau(x(t)) - \varphi_\tau(x^\tau, \alpha'_\tau))\beta_\tau(x(t))\right). \qquad (9)$$

Первое и второе слагаемые модели (9) отражают информацию, содержащуюся в точках обучающей выборки, а третье – во взаимосвязи между ними. Таким образом, в одном решающем правиле сочетаются возможности параметрических и локальных методов аппроксимации, что соответствует особенностям непараметрических моделей коллективного типа. Однако большее количество упрощенных аппроксимаций в предлагаемых моделях (5) создают реальную основу их преимущества над традиционными непараметрическими коллективами [2].

Для доказательства данного утверждения исследуем отношения асимптотических выражений среднеквадратических отклонений рассматриваемых непараметрических коллективов.

3. **Исследование асимптотических свойств непараметрических моделей коллективного типа.** Предварительно рассмотрим асимптотические свойства традиционного непараметрического коллектива (3), которые будут

использованы при анализе изучаемой статистики (5). Для упрощения выкладок будем считать, что закон распределения аргументов временной зависимости (1) является постоянным и известен, т. е. $p_t(x(t)) = p(x)\ \forall\ t \in T$, причем $x \in R^1$. В качестве упрощенных аппроксимаций примем линейные:

$$\varphi_t(x(t), \alpha_t) = \alpha_t x(t) + \alpha_t^0, \quad t = \overline{1, n-1}.$$

Найдем значение α_t^0 из условия $y^t = \alpha_t x^t + \alpha_t^0$:

$$\alpha_t^0 = y^t - \alpha_t x^t, \qquad \varphi_t(x(t), \alpha_t) = y^t + \alpha_t(x(t) - x^t).$$

Тогда традиционный непараметрический коллектив (3) примет вид

$$\bar{y}(t) = (p(x)(n-1)c)^{-1} \sum_{\tau=1}^{n-1} y^\tau \Phi\left(\frac{x(t) - x^\tau}{c}\right) +$$

$$+ (p(x)(n-1)c)^{-1} \sum_{\tau=1}^{n-1} \alpha^\tau (x(t) - x^\tau) \Phi\left(\frac{x(t) - x^\tau}{c}\right). \tag{10}$$

Без потери общности результатов будем считать α_τ не случайными параметрами и обозначим $x(t) = x$.

В этих условиях рассмотрим асимптотические свойства статистики (10).

Теорема 1. Пусть

1) восстанавливаемая зависимость $\psi_t(x)$ и плотности вероятности $p(x(t)) = p(x)$, $p(x(t), y(t)) = p(x, y)\ \forall\ t \in T$ ограничены со всеми своими производными до порядка 2 включительно;

2) ядерные функции $\Phi(u) \in H$;

3) последовательности $c(n) = c \to 0$, а $nc \to \infty$ при $n \to \infty$.

Тогда непараметрическая модель коллективного типа (10) обладает свойствами асимптотической несмещенности и состоятельности.

Доказательство. По определению

$$M(\bar{y}(t)) = \frac{1}{p(x)c} \int\int y \Phi\left(\frac{x-u}{c}\right) p(y, u) dy du +$$

$$+ \frac{\bar{\alpha}}{p(x)c} \int\int (x-u) \Phi\left(\frac{x-u}{c}\right) p(u) du, \tag{11}$$

где $\bar{\alpha} = \frac{1}{(n-1)} \sum_{\tau=1}^{n-1} \alpha_\tau$; M – знак математического ожидания.

После замены переменных $\left(\frac{x-u}{c}\right) = z$ и уравнения регрессии $\psi_t(u) = \int y p(y/u) dy = \psi(u)$ для каждого $t \in T$ получим выражение первого слагаемого (11):

$$\frac{1}{p(x)} \int \psi(x - cz) \Phi(z) p(x - cz) dz. \tag{12}$$

Разложим функции $\psi(x-cz)$, $p(x-cz)$ в ряд Тейлора в точке x и вычислим выражение (12) при $n \to \infty$:

$$\psi(x)+c^2\left[\frac{\psi(x)p^{(2)}(x)}{2p(x)}+\frac{\psi^{(1)}(x)p^{(1)}(x)}{p(x)}+\frac{\psi^2(x)}{2}\right]+O(c^4).$$

Следуя приведенной технологии преобразований, вычислим второе слагаемое выражения (11):

$$\frac{\overline{\alpha}}{p(x)}\int cz\Phi(z)p(x-cz)dz \sim -\frac{\overline{\alpha}p^{(1)}(x)}{p(x)}c^2+O(c^4).$$

Тогда при $n \to \infty$

$$\mathrm{M}(\overline{\psi}(x)-\psi(x)) \sim c^2\left[\frac{\psi(x)p^{(2)}(x)}{2p(x)}+\frac{\psi^{(1)}(x)p^{(1)}(x)}{p(x)}+\frac{\psi^2(x)}{2}-\frac{\overline{\alpha}p^{(1)}(x)}{p(x)}\right]+O(c^4), \tag{13}$$

что доказывает асимптотическую несмещенность традиционных непараметрических моделей коллективного типа, если при $n \to \infty$ коэффициент размытости ядерной функции $c(n) \to 0$.

Для доказательства второй части теоремы 1 определим асимптотическое выражение среднеквадратического отклонения при некотором $t \in T$:

$$\mathrm{M}(\overline{y}(t)-y(t))^2=\mathrm{M}(\overline{y}^2(t))-2y(t)\mathrm{M}(\overline{y}(t))+y^2(t). \tag{14}$$

Нетрудно показать, что

$$\mathrm{M}(\overline{y}^2(t))=\frac{1}{p^2(x)(n-1)^2c^2}\sum_{\tau=1}^{n-1}\sum_{\substack{\tau 1=1\\ \tau 1\neq\tau}}^{n-1}\mathrm{M}\left[(y^\tau+\alpha_\tau(x(t)-x^\tau))\Phi\left(\frac{x(t)-x^\tau}{c}\right)\times\right.$$
$$\left.\times(y^{\tau 1}+\alpha_{\tau 1}(x(t)-x^{\tau 1}))\Phi\left(\frac{x(t)-x^{\tau 1}}{c}\right)\right]+$$
$$+\sum_{\tau=1}^{n-1}\mathrm{M}\left((y^\tau+\alpha_\tau(x(t)-x^\tau))^2\Phi^2\left(\frac{x(t)-x^\tau}{c}\right)\right)=$$
$$=\frac{1}{p^2(x)c^2}\left[\left(\int(\psi(u)+\overline{\alpha}(x(t)-u))\Phi\left(\frac{x(t)-u}{c}\right)p(u)du\right)^2+\right.$$
$$+(n-1)^{-1}\int(\psi^2(u)+2\psi(u)\overline{\alpha}(x(t)-u)+\overline{\alpha}^2(x(t)-u)^2)\times$$
$$\left.\times\Phi^2\left(\frac{x(t)-u}{c}\right)p(u)du\right]. \tag{15}$$

Первое слагаемое выражения (15) с учетом сомножителя соответствует

$$(\mathrm{M}(\bar{y}(t)))^2 \sim \psi^2(x) + 2c^2\psi(x)A(x) + c^4A^2(x), \tag{16}$$

где

$$A(x) = \frac{\psi(x)p^{(2)}(x)}{2p(x)} + \frac{\psi^{(1)}(x)p^{(1)}(x)}{p(x)} + \frac{\psi^2(x)}{2} - \frac{\bar{\alpha}p^{(1)}(x)}{p(x)}.$$

Рассмотрим последовательно члены второго слагаемого:

$$\frac{1}{p^2(x)(n-1)c^2}\int \psi^2(x)\Phi^2\left(\frac{x(t)-u}{c}\right)p(u)du =$$

$$= \frac{1}{p^2(x)(n-1)c}\int \psi^2(x(t)-cz)p(x(t)-cz)\Phi^2(u)dz \sim$$

$$\sim \frac{\psi^2(x)}{(n-1)cp(x)} + O\left(\frac{c}{n-1}, c^4\right); \tag{17}$$

$$\frac{2\bar{\alpha}}{p^2(x)(n-1)c^2}\int \psi(u)(x(t)-u)\Phi^2\left(\frac{x(t)-u}{c}\right)p(u)du \sim$$

$$\sim \frac{2\bar{\alpha}}{p^2(x)(n-1)}\int \psi(x(t)-cz)z\Phi^2(z)p(x(t)-cz)dz \sim O\left(\frac{1}{n-1}\right);$$

$$\frac{\bar{\alpha}^2}{p^2(x)c^2(n-1)}\int (x(t)-u)^2\Phi^2\left(\frac{x(t)-u}{c}\right)p(u)du \sim O\left(\frac{c}{n-1}\right).$$

Подставляя асимптотические выражения (16) и (17) в выражение (14) с учетом $\mathrm{M}(\bar{y}(t))$ (13), получим

$$\mathrm{M}(\bar{y}(t)-y(t))^2 \sim \psi^2(x) - 2c^2\psi(x)A(x) + c^4A^2(x) + \frac{\psi^2(x)}{(n-1)cp(x)} -$$

$$-2\psi^2(x) - 2\psi(x)c^2A(x) + \psi^2(x) + O\left(\frac{1}{n-1}, \frac{c}{n-1}\right) =$$

$$= \frac{\psi^2(x)}{(n-1)cp(x)} + c^4A^2(x) + O\left(\frac{1}{n-1}, \frac{c}{n-1}\right). \tag{18}$$

Последнее выражение стремится к нулю, если с ростом $n \to \infty$ значения $c \to 0$, а $(n-1)c \to \infty$.

Из сходимости в среднеквадратическом (18) и свойства асимптотической несмещенности (13) следует свойство состоятельности непараметрического коллектива $\bar{y}(t)$.

Следствие теоремы 1. Найдем минимально возможное значение среднеквадратического отклонения (18). Для этого определим оптимальное значение параметра $c(n)$ и вычислим выражение (18).

Из условия минимума асимптотического выражения среднеквадратического отклонения (18) имеем оптимальное значение $c(n)$: $c^* = \left(\dfrac{\psi^2(x) \int \Phi^2(u)du}{(n-1)p(x)A(x)} \right)^{1/5}$, при котором

$$M(\bar{y}(t) - y(t))^2 = \frac{5}{4} \left[\frac{\psi^2(x) \int \Phi^2(u)du (A(x))^{1/2}}{(n-1)p(x)} \right]^{4/5}. \tag{19}$$

Основываясь на результатах и условиях теоремы 1, исследуем асимптотические свойства непараметрической модели коллективного типа (5).

Теорема 2. Пусть выполняются условия 1–3 теоремы 1. Тогда непараметрическая модель коллективного типа (5) обладает свойствами асимптотической несмещенности и состоятельности.

Доказательство. Проведем очевидные преобразования $M(\bar{\bar{y}}(t))$, считая $\bar{y}'(t) = \bar{\psi}_t'(x(t))$, $\bar{y}(t) = \bar{\psi}_t(x(t))$:

$$M(\bar{\bar{y}}(t) - y(t)) = M(\beta \bar{y}'(t) + (1-\beta)\bar{y}(t) - y(t)) = M(\bar{y}(t) - y) +$$

$$+ \beta[M(\bar{y}'(t) - y) - M(\bar{y}(t) - y)] \sim c^2 \left[A(x) + \beta \frac{p^{(1)}(x)}{p(x)} (\bar{\alpha}' - \bar{\alpha}) \right], \tag{20}$$

что подтверждает асимптотическую несмещенность предлагаемой непараметрической модели $\bar{\bar{y}}(t)$.

Вычислим

$$M(\bar{\bar{y}}(t) - y(t))^2 = (1+\beta^2) M(\bar{y}(t) - y(t))^2 + \beta^2 M(\bar{y}'(t) - y(t))^2 -$$

$$-2\beta^2 M(\bar{y}'(t) - y(t)) M(\bar{y}(t) - y(t)) + 2\beta M((\bar{y}(t) - y(t))(\bar{y}'(t) - y(t))) -$$

$$-2\beta M(\bar{y}(t) - y(t))^2 = (1-\beta)^2 M(\bar{y}(t) - y(t))^2 + \beta^2 M(\bar{y}'(t) - y(t))^2 +$$

$$+2\beta(1-\beta) M((\bar{y}(t) - y(t))(\bar{y}'(t) - y(t))) \le$$

$$\le [(1-\beta)(M(\bar{y}(t) - y(t))^2)^{1/2} + \beta(M(\bar{y}'(t) - y(t))^2)^{1/2}]^2. \tag{21}$$

Отсюда следует сходимость в среднеквадратическом и состоятельность непараметрических моделей, основанных на методе двойного коллективного оценивания.

4. **Сравнение непараметрических моделей коллективного типа.** Исследуем отношение асимптотических выражений среднеквадратических критериев для моделей (3) и (5):

$$W = \frac{M(\bar{\bar{y}}(t) - y)^2}{M(\bar{y}(t) - y)^2} \tag{22}$$

и определим условия выполнения неравенства $W < 1$, которые соответствуют области компетентности $\bar{\bar{y}}(t)$.

С этой целью воспользуемся следствием теоремы 1 и оценим в (22) минимальные значения составляющих при оптимальных коэффициентах размытости.

Получим

$$\mathrm{M}(\bar{\bar{y}}(t)-y)^2 = \frac{5}{4}\left(\frac{\psi^2(x)\int\Phi^2(u)du}{(n-1)p(x)}\right)^{4/5}[(1-\beta)(A(x))^{1/4}+\beta(A'(x))^{1/4}]^2, \quad (23)$$

где

$$A'(x) = \frac{\psi(x)p^{(2)}(x)}{2p(x)} + \frac{\psi^{(1)}(x)p^{(1)}(x)}{p(x)} + \frac{\psi^2(x)}{2} - \frac{\bar{\alpha}'p^{(1)}(x)}{p(x)},$$

$\bar{\alpha}' = \sum_{\tau=1}^{n}\alpha'_\tau/(n-1)$ – среднее значение коэффициентов для модели $\bar{y}'(t)$.

Тогда с учетом (19), (23) отношение (22) представляется в виде

$$W = \left[\frac{(1-\beta)A(x)+\beta A'(x)}{A(x)}\right]^{2/5} = \left[(1-\beta)+\beta\frac{A'(x)}{A(x)}\right]^{2/5}.$$

Выполнение соотношения $W < 1$ соблюдается, если $1-\beta+\beta\frac{A'(x)}{A(x)} < 1$ либо $p^{(1)}(x)(\bar{\alpha}-\bar{\alpha}') < 0$.

Если $p^{(1)}(x) > 0$, то непараметрическая модель с двойным коллективным оцениванием имеет преимущество над традиционными коллективами (3) при $\bar{\alpha}' > \bar{\alpha}$ в условиях $\bar{\alpha}'$, $\bar{\alpha} > 0$ и при $|\bar{\alpha}'| < |\bar{\alpha}|$, если $\bar{\alpha}'$, $\bar{\alpha} < 0$.

Если $p^{(1)}(x) < 0$, область компетентности определяется условиями: $\bar{\alpha} > \bar{\alpha}'\ \forall\ \bar{\alpha}, \bar{\alpha}' > 0$; $|\bar{\alpha}'| > |\bar{\alpha}|\ \forall\ \bar{\alpha}', \bar{\alpha} < 0$.

Наличие различных знаков $\bar{\alpha}$ и $\bar{\alpha}'$ является критерием непредставительности исходных данных.

При равномерном законе распределения $p(x)$ непараметрические модели коллективного типа имеют равную эффективность, так как введение дополнительных упрощенных аппроксимаций не привносит новую информацию. С вычислительной точки зрения целесообразно применять традиционный коллектив.

Критерием применения двойного коллективного оценивания является достоверное отличие $\bar{\alpha}, \bar{\alpha}'$ в конкретных условиях $x(t)$.

Естественно желание объяснить физическую причину полученных условий преимуществами предложенных непараметрических коллективов. Например, при $p^{(1)}(x) > 0$ в области больших значений восстанавливаемой зависимости ($\bar{\alpha}, \bar{\alpha}' > 0$) наблюдается большее накопление экспериментальных данных, чем левее точки x. В этом случае более целесообразно формирование упрощенных аппроксимаций «слева направо» (от меньших значений x к большим), так как это позволяет более точно оценить их параметры по большему количеству наблюдений.

Заключение. Исследование уникальных экологических и социально-экономических систем связано с решением проблем анализа коротких временных рядов наблюдений их параметров, когда отношение объем выборки/размерность невелико. В данных условиях необходимо проведение управляемой предварительной обработки исходной информации с целью построения частных моделей изучаемой временной зависимости и их обобщение в едином решающем правиле, что реализуется в предлагаемом подходе. Показана возможность увеличения объема интегрированных сведений в виде упрощенных аппроксимаций и установлена эффективность их применения при моделировании временных процессов на основе принципа двойного коллективного оценивания.

Перспективным является развитие исследований, ориентированных на проблему моделирования комплекса взаимосвязанных временных процессов с учетом их нестационарного характера.

СПИСОК ЛИТЕРАТУРЫ

1. **Лапко А. В., Ченцов С. В., Крохов С. И., Фельдман Л. А.** Обучающиеся системы обработки информации и принятия решений. Новосибирск: Наука, 1996.
2. **Крившич Д. В., Ченцов С. В.** Коллективы непараметрических моделей нестационарных временных зависимостей // Вестник Красноярского государственного технического университета. Красноярск: КГТУ, 1996. Вып. 5. С. 90.

Институт вычислительного моделирования СО РАН,
E-mail: lapko@ksc.krasn.ru

Поступила в редакцию
1 декабря 2000 г.