РОБАСТНАЯ ИДЕНТИФИКАЦИЯ НЕЛИНЕЙНЫХ ОБЪЕКТОВ

УДК 519.71

РУДЕНКО Олег Григорьевич
д.т.н., проф., зав. каф. ЭВМ, Харьковский национальный университет радиоэлектроники
Научные интересы: искусственный интеллект, интеллектуальные системы
E-mail: rudenko@kharkov.kture.ua

БЕССОНОВ Александр Александрович
к.т.н., доцент, кафедра ЭВМ, Харьковский национальный университет радиоэлектроники
Научные интересы: искусственные нейронные сети, искусственные иммунные системы
E-mail: o.bezsonov@gmail.com

РУДЕНКО Сергей Олегович
аспирант, кафедра ЭВМ, Харьковский национальный университет радиоэлектроники
Научные интересы: искусственный интеллект, интеллектуальные системы
E-mail: rudenko@kharkov.kture.ua

Введение. Задача идентификации нелинейного нестационарного динамического объекта, описываемого моделью

$$y(k) = f(x(k),k) + \xi(k)\,, \qquad (1)$$

где

$x(k) = \left[y(k-1),..., y(k-l), u(k-1),..., u(k-n)\right]^T$ - Nx1 вектор обобщенного входного сигнала ($N = l + n$); $y(i)$, $u(i)$ - выходной и входной сигналы объекта в момент времени i соответственно; l и n – порядки запаздывания по выходному и входному каналам соответственно; $f(\bullet)$ - неизвестная нелинейная функция; $\xi(k)$ - помеха, состоит в получении оценки функции $f(\bullet)$ по измерениям входных и выходных переменных.

Классические методы построения модели (1) используют аппроксимацию нелинейности $f(\bullet)$ рядами или полиномами и требуют решения задач структурной (определение вида и степени аппроксимирующего полинома и т.д.) и параметрической (определение коэффициентов разложения) идентификации. Сложность получения адекватного математического описания существенно возрастает, если параметры исследуемого объекта или условия его функционирования являются нестационарными. Также в настоящее время в основе подавляющего большинства методов идентификации лежит гипотеза о нормальном распределении наблюдаемых величин. Следствием этого является широкое использование критерия минимума суммы квадратов ошибок наблюдений и связанного с ним метода наименьших квадратов. Опыт, однако, показывает, что методы идентификации, основанные на МНК, оказываются чрезвычайно чувствительными к отклонениям фактического закона распределения от нормального. В условиях различного рода сбоев, выбросов, помех, распределенных по законам, отличным от нормального, метод наименьших квадратов теряет работоспособность. Тем не менее, и сегодня метод наименьших квадратов является основным при решении задач построения математических моделей, хотя его корректное использование ограничивается целым рядом обстоятельств. В практических задачах малый объем информации в большинстве случаев не позволяет надежно установить вид закона распределения, в том же случае, если истинное распределение имеет «тяжелые хвосты», например, распределение Лапласа, Коши, эффективность

МНК резко падает. Процедуры, связанные с МНК, крайне чувствительны к выбросам в данных и грубым ошибкам, которые, во-первых, достаточно трудно выявить, а во-вторых, разделение наблюдений на нормальные и грубо ошибочные довольно искусственно и не имеет смысла для медленно убывающих распределений. Кроме того нет никаких оснований, кроме математического удобства, использовать линейность или несмещенность оценок.

Альтернативный метод, который направлен на устранение влияния различного рода помех заключается в замене принципа наименьших квадратов принципом минимизации суммы абсолютных величин (модулей) отклонений. Таким образом, можно получить оценки более устойчивые по отношению к сильным отклонениям от нормальности, причем наиболее эффективные для распределения Лапласа. Необходимо, однако, отметить, что этим оценкам присущ ряд собственных недостатков, таких как неоднозначность решения задачи минимизации суммы модулей ошибок, определенные вычислительные трудности и другие.

В связи с этим для решения задачи идентификации нелинейных объектов в последнее время все более широко применяются искусственные нейронные сети (ИНС).

Среди ИНС наибольшее распространение при решении задачи идентификации нелинейных динамических объектов (1) в настоящее время получили многослойный персептрон (МП) и радиально-базисные сети (РБС), использующие соответственно аппроксимации нелинейного оператора *f(•)* вида

$$\hat{y}(k) = \hat{f}(k) =$$
$$f^{q}\left[\left(W^{q}\right)^{T} f^{q-1}\left[\left(W^{q-1}\right)^{T} f^{q-2}\left[\ldots f^{1}\left[\left(W^{1}x(k)+b_{1}\right)^{T}\right]\ldots\right]\right]\right]+ \quad (2)$$
$$+b_{q},$$

и

$$\hat{y}(k) = \hat{f}(k) = w_0 + \sum_{i=1}^{N} w_i \Phi_i(x) = w_0 + W^T \Phi(k)\,, \quad (3)$$

где W^i – вектор весовых параметров нейронов *i*-го слоя сети; $f^i[\bullet]$ – активационная функция (АФ) *i*-го слоя, b_i – смещение *i*-го нейрона; $\Phi(k)$ – вектор выбранных базисных функций (БФ).

При выборе критерия

$$F(\theta) = \sum_{i=1}^{K} \rho\left(e(i,\theta)\right), \quad (4)$$

обучение ИНС сводится к поиску оценки $\hat{\theta}$, определяемой как решение системы уравнений

$$\nabla F\left(\theta_j\right) = \frac{\partial F(\theta)}{\partial \theta_j} = \sum_{i=1}^{K} \psi\left(e(i,\theta)\right) \frac{\partial e(i,\theta)}{\partial \theta_j} = 0, \quad (5)$$

где $\rho\left(e(i,\theta)\right)$ - некоторая функция потерь; $e(i,\theta) = y(i) - \hat{y}(i,\theta)$; $\psi\left(e(i,\theta)\right) = \frac{\partial \rho\left(e(i,\theta)\right)}{\partial e(i,\theta)}$ - функция влияния.

При этом в качестве оценок коэффициентов нейросетевой модели могут использоваться:

1. Оценки Хубера, получаемые минимизацией критерия

$$Q = \sum_{i=1}^{n} \rho\left(\mathrm{y}[\mathrm{i}] - \sum_{\mathrm{j}=1}^{\rho} x_j[i]\beta_{\mathrm{i}}\right) = \sum_{i=1}^{n} \rho\left(\varepsilon[i]\right), \quad (6)$$

где $\varepsilon[\mathrm{i}]$ – ошибка i–го наблюдения, а ρ имеет вид

$$\rho(\varepsilon) = \begin{cases} \frac{1}{2}\varepsilon^2, & |\varepsilon| \le c \\ \mathrm{c}\,|\varepsilon| - \frac{1}{2}\mathrm{c}^2, & |\varepsilon| > c \end{cases}. \quad (7)$$

При этом полагаем, что с = *kS*, где *k* обычно равно 1.5 или 2.1, а *S* – оценка параметра масштаба.

2. Оценки Эндрюса. При этом наблюдениям, ошибки которых велики по абсолютной величине, присваиваются нулевые веса:

$$\psi(\varepsilon) = \begin{cases} 2.1\mathrm{S}\left(1 - \cos\left(\frac{\varepsilon}{2} 1\mathrm{S}\right)\right), & |\varepsilon| \le 2.1S\pi \\ 4.2\mathrm{S}, & |\varepsilon| > 2.1S\pi \end{cases}. \quad (8)$$

3. Оценки Мэллоуза являются модификацией хуберовский оценок, в которых наряду со взвешиванием остатков ε производится взвешивание факторов, что позволяет уменьшить влияние точек, резко выделяющихся в пространстве независимых переменных. Для нахождения оценок Мэллоуза необходимо решить $\rho+1$ уравнений

$$\psi(\varepsilon) = \begin{cases} \sum_{i=1}^{n} \psi(\varepsilon[\mathrm{i}]) \\ \sum_{i=1}^{n} x_m[i]\psi^{x}\left(x[\mathrm{i}]\right)\psi\left(\varepsilon[\mathrm{i}]\right), \ \mathrm{m} = 0,1,..,\rho \end{cases}, \quad (9)$$

где ψ определяется по Хуберу, а

$$\psi^{x}\left(x[\mathrm{i}]\right) = \prod_{j=1}^{\rho} \psi^{(j)}\left(x[\mathrm{i}]\right). \quad (10)$$

4. В оценках Форсайта минимизирующий критерий имеет вид

$$Q = \sum_{i=1}^{n} \rho(\varepsilon[\mathrm{i}]) = \sum_{i=1}^{n} |\,\varepsilon[\mathrm{i}]\,|^{\alpha}, \qquad (11)$$

где $1 \le \alpha \le 2$. Эти оценки близки к хуберовским, однако, не имеют столь убедительного теоретического обоснования. Эмпирически показано, что приемлемым является $\alpha = 1.5$. При $\alpha = 2$ оценки Форсайта совпадают с оценками МНК, а при $\alpha = 1$ получаем метод наименьших абсолютных отклонений (модулей), при котором критерий идентификации имеет вид

$$Q = \sum_{i=1}^{n} |\varepsilon[\mathrm{i}]|. \qquad (12)$$

5. Для получения оценок Меррила-Швеппа используется

$$\rho(\varepsilon) = \begin{cases} \gamma_1 \varepsilon^2, |\varepsilon| \le k \\ \gamma_2 \sqrt{\varepsilon}, |\varepsilon| > k \end{cases}, \qquad (13)$$

где $\gamma_1 > 0, \gamma_2 > 0, 0 \le \mathrm{k}$.

В случае нормального распределения, оптимальные оценки, получаемые с помощью метода наименьших квадратов (МНК), инвариантны к параметру масштаба σ_ξ^2. Если же помехи имеют распределения, отличные от нормального, МНК-оценки оказываются неустойчивыми и необходимо применять неквадратичные функции потерь $\rho(e(i,\theta))$. При М-оценивании, вследствие неоднородности функции ρ, получаемые оценки не сохраняют свойства инвариантности. Для того чтобы свойство инвариантности масштаба выполнялось, в функционале (5) вместо ошибки $e(i,\theta)$ следует брать $\tilde{e}(i,\theta) = (e(i,\theta) - m)/S$, где S - помехоустойчивая оценка параметра масштаба или мера рассеяния остаточных разностей (в случае нормального распределения S является оценкой σ_ξ). От того, насколько удачно оценен параметр масштаба, зависит эффективность робастного обучения сети.

При решении задач в режиме on-line естественным является рекуррентное оценивание параметра масштаба. Так в [1] в качестве рекуррентной рассматривалась следующая оценка:

$$S_n = \beta S_{n-1} + (1-\beta)\min\left[3S_{n-1}, |e(n)|\right], \qquad (14)$$

где $\beta = 0.75$.

Наличие в данном алгоритме второго слагаемого обеспечивает робастность получаемой оценки к выбросам.

Градиентный алгоритм оценивания параметра S

$$S_n = \lambda S_{n-1} + \frac{1-\lambda}{\beta} S_{n-1} \psi\left(\frac{|e(n)|}{S_{n-1}}\right) \qquad (15)$$

в предположении, что $\psi(\bullet)$ является ограниченной функцией

$$\psi\left(\frac{|e(n)|}{S}\right) = \min\left\{\frac{|e(n)|}{S}, k_0\right\},$$

рассматривался в [2]. В выражении (15) $\lambda \in [0.99,1]$, β - константа, зависящая от k_0, выбираемая из интервала $[0.6, 0.74]$.

Следует отметить, что достаточно общей моделью засорения является модель Тьюки-Хьюбера [3]

$$\rho(x) = (1-\varepsilon)\rho_0(x) + \varepsilon q(x), \qquad (16)$$

где $\rho_0(x)$ - плотность соответствующего основного распределения; $q(x)$ - плотность засоряющего (произвольного распределения); $\varepsilon \in [0,1]$ - параметр, характеризующий степень засорения основного распределения.

При использовании модели засорения (16) необходимо также оценить величины S_1^2 и S_2^2 и учесть эти оценки в алгоритме обучения. Если σ_1^2 и σ_2^2 не изменяются во времени, такое оценивание также может быть осуществлено методом стохастической аппроксимации:

$$S_{1,n}^2 = \begin{cases} S_{1,n-1}^2 + \dfrac{1}{l_{1,n}}\left(\tilde{e}_n^2 - S_{1,n-1}^2\right), \ е\ есл\left|\tilde{e}_n\right| \le 3S_{1,n-1}, \\ S_{1,n-1}^2\ \textit{в противном случае}, \end{cases}$$

$$S_{2,n}^2 = \begin{cases} S_{2,n-1}^2 + \dfrac{1}{l_{2,n}}\left(\tilde{e}_n^2 - S_{2,n-1}^2\right), \ е\ есл\left|\tilde{e}_n\right| > 3S_{1,n-1}, \\ S_{2,n-1}^2\ \textit{в противном случае}; \end{cases}$$

где

$$l_{1,n} = n - l_{2,n};$$

$$l_{2,n} = \begin{cases} 0,\ е\ есл\left|\tilde{e}_n\right| \le 3S_{1,n-1}, \\ l_{2,n\text{-}1} + 1\ \textit{в противном случае}. \end{cases}$$

Общая дисперсия помехи вычисляется по формуле

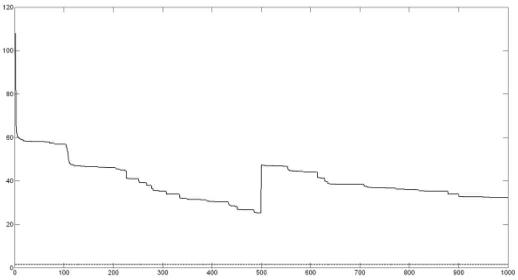
$$S_n^2 = \begin{cases} S_{1,n}^2,\ е\ есл \left|\tilde{e}_n\right| \le 3S_{1,n-1}; \\ S_{2,n}^2\ в\ противном\ случае \end{cases}$$

и используется как параметр масштаба в квадратичном функционале [4].

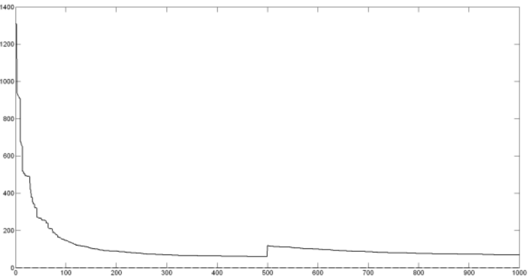
Моделирование. Решалась задача идентификации сильно зашумленного объекта

$$y(k) = 0.725\beta(k)$$
$$\sin\left(\frac{16u(k-1)+8y(k-1)}{\beta(k)\left(3+4u^2(k-1)+4y^2(k-1)\right)}\right)+$$
$$+0.2u(k-1)+0.2y(k-1)+\xi(k),$$

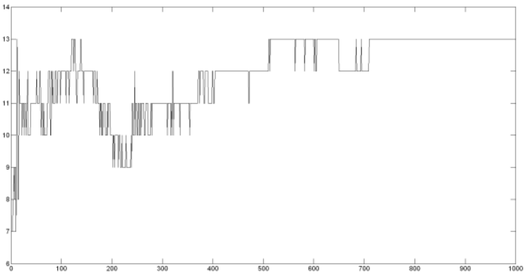
где $\beta(k)=1$, $\xi(k)=(1-\varepsilon)q_1(k)+\varepsilon q_2(k)$ - засоренная помеха ($q_1(k)$, $q_2(k)$ – нормально распределенные помехи с дисперсиями σ_1^2 и σ_2^2 соответственно); с использованием МП и РБС. На 500-ом шаге обучения происходило изменение параметров помехи с $\sigma_2=6$ на $\sigma_2=12$. Результаты идентификации приведены на рис.1. На рис. 1-а), б) приведены графики изменения ошибок обучения сетей МП и РБС соответственно, а на рис. 1-в) показан график изменения количества нейронов шаблонного слоя сети РБС в ходе обучения.



а)



б)



в)

Рис. 1 – Результаты идентификации объекта (21) с помощью МП и РБС

ЗАКЛЮЧЕНИЕ

Эффективность использования нейросетевого подхода для решения задачи идентификации нелинейных объектов при наличии негауссовских помех существенно возрастает при применении робастных функционалов обучения. При этом возникает необходимость оценивания используемого параметра масштаба. При on-line обучении предпочтение следует отдавать рекуррентному оцениванию. Результаты имитационного моделирования подтверждают эффективность предложенного подхода.

ЛИТЕРАТУРА:

1. Deng G. Sequential and adaptive learning algorithms for M-estimation // EURASIP J. Advances in Signal Processing. – 2008. – ID459586. – 13 p.
2. Gänsler T., Gay S.L., Sondni M.M., Benesty J. Double-talk robust fast converging algorithms for network echo cancellation // IEEE Tr. Speech and Audio Processing. – 2000. – v.8. – №6. – Pp. 656-663.
3. Hogg R.V. Vvedenie v pomehoustojchivoe ocenivanie // V kn. «Ustojchivye stohasticheskie metody ocenki dannyh» - M.: Mashinostroenie, 1984. – S.12-25.
4. Bessonov A.A., Rudenko O.G., Rudenko S.O. Ustojchivoe obuchenie radial'no-bazisnyh setej s ogranichennoj tochnost'ju / Vestnik HNTU. – 2010. – 2(38). – S.130-134.