

Е.В. Бодянский, Е.А. Винокурова, И.О. Кобылин, П.П. Мулеса

## Робастная адаптивная идентификация нестационарных временных рядов с помощью ансамбля обучаемых гибридных адаптивных моделей

Предложены модели робастной адаптивной идентификации нестационарных временных рядов в *on-line* режиме поступления потока данных, характеризующиеся простотой вычислительной реализации. Проведен ряд имитационных экспериментов на основе тестовых и реальных данных.

**Ключевые слова:** адаптивная идентификация, робастные целевые функции, нестационарные нелинейные временные ряды, ансамбль моделей.

Запропоновано моделі робастної адаптивної ідентифікації нестационарних часових рядів в *on-line* режимі надходження потоку даних, що характеризуються простотою обчислювальною реалізацією. Проведено низку імітаційних експериментів на основі тестових і реальних даних.

**Ключові слова:** адаптивна ідентифікація, робастні цільові функції, нестационарні нелінійні часові ряди, ансамбль моделей.

**Введение.** Широкий класс задач интеллектуального анализа потоков данных (*Data Stream Mining*) связан с исследованием и обработкой временных рядов различной природы, при этом, как правило, эти задачи включают в себя идентификацию, прогнозирование, обнаружение разладок, сегментацию, кластеризацию и т.д. [1, 2]. Обычно предполагается, что обрабатываемые последовательности заданы заранее в форме пакета наблюдений, чьи свойства не меняются с течением времени.

Ситуация существенно усложняется, когда данные на обработку поступают последовательно в *on-line* режиме в форме потока данных [3, 4], при этом свойства рядов могут нестационарным образом изменяться во времени.

В этой ситуации на первый план выходит адаптивный подход [5], связанный с использованием обучаемых моделей, настройка которых проводится в реальном времени с помощью тех или иных рекуррентных процедур.

Таким образом, актуален синтез адаптивных моделей, которые могли бы обрабатывать нестационарные временные потоки данных, в том числе в ситуациях, когда анализируемый временной ряд загрязнен аномальными выбросами.

### Постановка задачи

Оценивание параметров стохастических объектов в условиях неопределенности требует наличия эффективных алгоритмов идентификации, обеспечивающих синтез математических моделей по данным наблюдений за входными и вы-

ходными сигналами исследуемого явления. Широкий класс временных рядов может быть описан с помощью уравнения псевдолинейной регрессии [6].

$$y(k) = w^T x(k) + \zeta(k), \quad (1)$$

где  $y(k)$  – скалярный выход объекта (отклик) в дискретный момент времени  $k = 1, 2, \dots, N$ ,  $w = (w_0, w_1, \dots, w_n)^T - ((n+1) \times 1)$  – вектор неизвестных параметров, подлежащих определению,  $x(k) \in R^n$  – вектор входных переменных (факторов, регрессоров),  $\zeta(k)$  – случайное возмущение (помеха) с нулевым математическим ожиданием и неизвестной функцией плотности распределения. В рамках уравнения (1) могут быть описаны популярные модели временных рядов, такие как авторегрессионная (AR) модель, модель скользящего среднего (MA), авторегрессионная модель скользящего среднего (ARMA), авторегрессионная интегрированная расширенная модель скользящего среднего (ARIMAX) и т.д., связанные с концепцией фильтра Бокса–Дженкинса.

В основе значительного числа процедур и алгоритмов синтеза таких моделей лежит гипотеза о нормальном распределении помех –  $\zeta(k) \sim N(0, \sigma^2)$ , что привело к повсеместному использованию критерия минимума суммы квадратов ошибок оценивания и связанному с ним методу наименьших квадратов в различных модификациях. В случае если данные на обработку поступают последовательно либо объем выборки  $N$  не фиксирован, предпочтение следует отдать рекуррентному методу наимень-

ших квадратов, который, однако, при больших значениях  $N$  может быть численно неустойчив, громоздок и в итоге приводит к *взрыву параметров* ковариационной матрицы.

### Адаптивная модель нестационарного ряда на основе квадратичного критерия

Эффективной альтернативой рекуррентному методу наименьших квадратов есть специализированная процедура стохастической аппроксимации Гудвина–Ремеджа–Кейнеса для идентификации объектов управления, предложенная в [7]. Рассматривая уравнение настраиваемой модели

$$\hat{y}(k) = \hat{w}^T(k)x(k), \quad (2)$$

где  $\hat{w}(k) - ((n+1) \times 1)$  – вектор настраиваемых параметров, ошибку идентификации

$$e(k) = y(k) - \hat{y}(k) = y(k) - \hat{w}^T(k-1)x(k), \quad (3)$$

функцию потерь

$$\rho(e(k)) = \frac{1}{2} e^2(k) \quad (4)$$

и основанный на функции потерь (4) критерий идентификации

$$E(e(k)) = \sum_k \rho(e(k)) = \frac{1}{2} \sum_k e^2(k), \quad (5)$$

можно записать алгоритм адаптивной идентификации в виде системы рекуррентных соотношений

$$\begin{cases} \hat{w}(k) = \hat{w}(k-1) + \frac{(y(k) - \hat{w}^T(k-1)x(k))x(k)}{r(k)}, \\ r(k) = r(k-1) + \|x(k)\|^2, \quad r(0) = 1. \end{cases} \quad (6)$$

При необходимости идентификации сигнала, параметры которого изменяются во времени, может быть использована модификация процедуры (6), основанная на критерии

$$E(e(k)) = \sum_k \alpha^{k-1} \rho(e(k)) = \frac{1}{2} \sum_k \alpha^{k-1} e^2(k), \quad (7)$$

где  $0 \leq \alpha \leq 1$  – параметр забывания информации.

При этом модифицированный алгоритм имеет вид:

$$\begin{cases} \hat{w}(k) = \hat{w}(k-1) + \frac{(y(k) - \hat{w}^T(k-1)x(k))x(k)}{r(k)}, \\ r(k) = \alpha r(k-1) + \|x(k)\|^2, \quad 0 \leq \alpha \leq 1. \end{cases} \quad (8)$$

Несложно заметить, что процедура (8) в отличие от экспоненциально взвешенного рекуррентного метода наименьших квадратов устойчива при любых значениях параметра забывания  $\alpha$ , совпадает при  $\alpha = 1$  с алгоритмом Гудвина–Ремеджа–Кэйнеса, а при  $\alpha = 0$  принимает форму популярного в теории обучения искусственных нейронных сетей алгоритма Качмажа–Уидроу–Хоффа.

На рис. 1 приведена структурная схема обучаемой модели, настраиваемой с помощью процедуры (5).

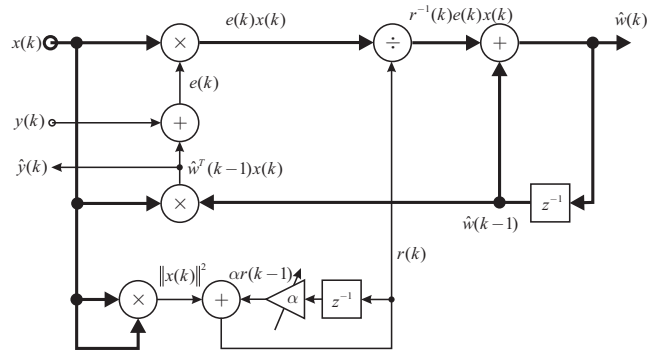


Рис. 1. Структурная схема обучаемой модели, настраиваемой с помощью алгоритма Гудвина–Ремеджа–Кэйнеса

Видно, что процесс оценивания реализуется с помощью элементарных арифметических операций и операций сдвига назад  $z^{-1}$ .

### Робастные адаптивные модели временных рядов

Многие реальные задачи обработки информации свидетельствуют о том, что распределение данных не так уж часто описывается гауссовским законом, хотя большинство стандартных статистических методов базируется на явных или неявных предположениях о нормальности закона распределения анализируемых рядов. В то же время известно, что методы идентификации, основанные на критерии наименьших квадратов, чрезвычайно чувствительны к отклонениям фактического закона распределения от нормального. В условиях различного рода сбоев, выбросов, грубых ошибок, негауссовских помех с *тяжелыми хвостами* метод наименьших квадратов теряет свою эффективность.

Именно этот факт привел к созданию широкого класса методов робастного оценивания, основанных на минимизации критериев, отличных

от квадратичного, и необходимости решения задачи нелинейной оптимизации. Поскольку при работе в *on-line* режиме решение этой задачи затруднено, в силу численной громоздкости, эффективных адаптивных методов робастной идентификации в полном смысле этого термина известно не так уж и много [8–10].

Сегодня известно достаточно много робастных функций потерь, общей чертой которых есть *подавление* далеко находящихся от точки экстремума наблюдений (выбросов). Весьма интересна функция *Geman–McClure*, имеющая вид [11–13]

$$\rho(e(k)) = \frac{1}{2} \cdot \frac{e^2(k)}{1 + e^2(k)}. \quad (9)$$

Интересно в этой функции то, что она порождена функцией плотности распределения Коши

$$\rho_c(e(k)) = \frac{1}{1 + e^2(k)} \quad (10)$$

и представляет собой по сути *перевернутый* и сдвинутый на единицу кошиан

$$\rho(e(k)) = 1 - \rho_c(e(k)). \quad (11)$$

Вводя в (10) масштабирующий параметр  $\sigma^2$ , определяющий *ширину* кошиана

$$\rho_c(e(k)) = \frac{1}{1 + \frac{e^2(k)}{\sigma^2}} = \frac{\sigma^2}{\sigma^2 + e^2(k)}, \quad (12)$$

можно получить модифицированную функцию *Geman–McClure*:

$$\rho(e(k)) = \frac{1}{2} \cdot \frac{e^2(k)}{\sigma^2 + e^2(k)}, \quad (13)$$

вид которой, при различных параметрах ширины  $\sigma^2$ , приведен на рис. 2.

Несложно также записать функцию влияния и весовую функцию для (13) в виде

$$\psi(e(k)) = \rho'(e(k)) = \frac{e(k)}{(\sigma^2 + e^2(k))^2}, \quad (14)$$

$$w(e(k)) = \frac{\psi(e(k))}{e(k)} = \frac{1}{(\sigma^2 + e^2(k))^2}. \quad (15)$$

Вводя далее целевую функцию

$$E(e(k)) = \sum_k \rho(e(k)) = \frac{1}{2} \sum_k \frac{e^2(k)}{\sigma^2 + e^2(k)}, \quad (16)$$

можно ввести градиентный алгоритм ее оптимизации в виде

$$\begin{aligned} \hat{w}(k) &= \hat{w}(k-1) + \eta(k) \psi(e(k)) x(k) = \\ &= \hat{w}(k-1) + \eta(k) e(k) \frac{\sigma^2 x(k)}{(\sigma^2 + e^2(k))^2} = \\ &= \hat{w}(k-1) + \eta(k) e(k) J_G(k), \end{aligned} \quad (17)$$

где  $\eta(k)$  – параметр шага обучения.

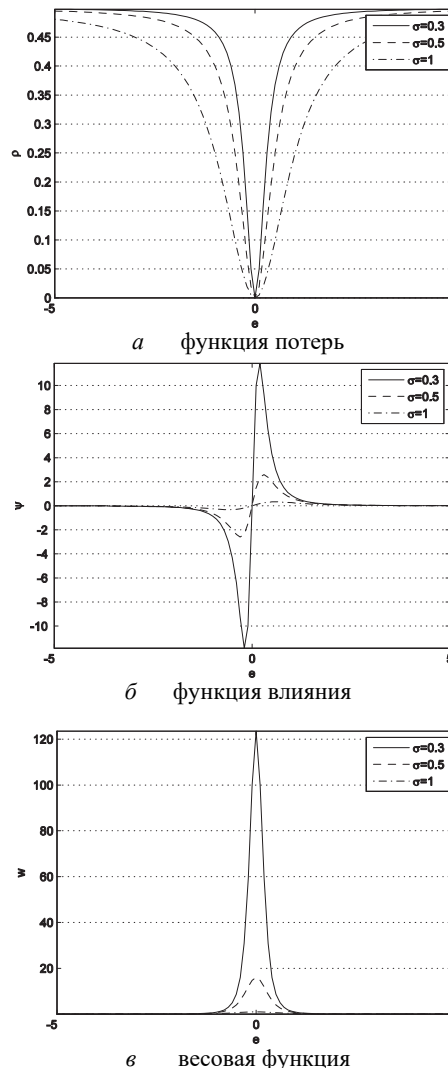


Рис. 2. Модифицированная функция *Geman–McClure*

Используя технику оптимизации алгоритмов обучения [15], можно ввести экспоненциально взвешенную процедуру типа (8), в виде

$$\begin{cases} \hat{w}(k) = \hat{w}(k-1) + \frac{(y(k) - \hat{w}^T(k-1)x(k)) J_G(k)}{r(k)} \\ r(k) = \alpha r(k-1) + \|J_G(k)\|^2, \quad 0 \leq \alpha \leq 1, \end{cases} \quad (18)$$

где  $J_G(k) = \frac{\sigma^2 x(k)}{(\sigma^2 + e^2(k))^2}$ .

На рис. 3 приведена структурная схема настраиваемой модели, обучаемой с помощью процедуры (18).

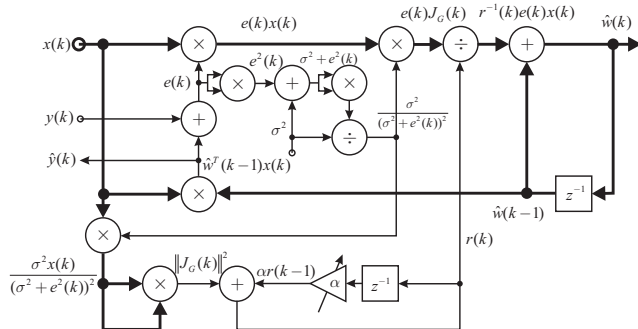


Рис. 3. Настраиваемая модель, обучаемая на основе функции потерь *Geman-McClure*

Отметим, что эта схема не намного сложнее системы, приведенной на рис. 1 и также реализуется с помощью элементарных операций.

Весьма близка по свойствам к (13) является робастная функция потерь Коши [13, 14, 16]

$$\rho(e(k)) = \frac{\sigma^2}{2} \ln \left( 1 + \frac{e^2(k)}{\sigma^2} \right), \quad (19)$$

вид которой при разных значениях  $\sigma^2$  приведен на рис. 4.

Несложно увидеть, что функция влияния

$$\psi(e(k)) = \frac{e(k)\sigma^2}{\sigma^2 + e^2(k)} \quad (20)$$

и весовая функция

$$w(e(k)) = \frac{\sigma^2}{\sigma^2 + e^2(k)} \quad (21)$$

близки  $\rho_c(e(k))$  к рассмотренным нами.

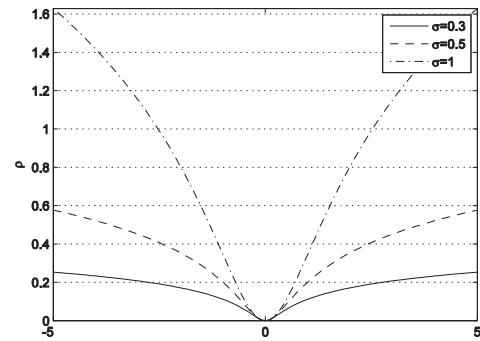
Следовательно, несложно записать градиентную процедуру

$$\begin{aligned} \hat{w}(k) &= \hat{w}(k-1) + \eta(k)\psi(e(k))x(k) = \\ &= \hat{w}(k-1) + \eta(k)e(k) \frac{\sigma^2}{\sigma^2 + e^2(k)} = \\ &= \hat{w}(k-1) + \eta(k)e(k)J_C(k) \end{aligned} \quad (22)$$

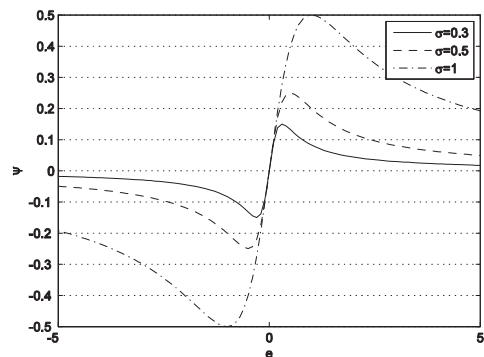
и аналог алгоритма (18) в виде

$$\begin{cases} \hat{w}(k) = \hat{w}(k-1) + \frac{e(k)J_C(k)}{r(k)}, \\ r(k) = \alpha r(k-1) + \|J_C(k)\|^2, \end{cases} \quad (23)$$

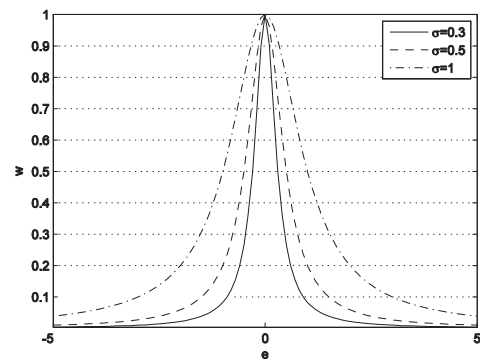
где  $J_C(k) = \frac{\sigma^2}{\sigma^2 + e^2(k)}$ .



а функция потерь



б функция влияния



в весовая функция

Рис. 4. Робастная функция потерь Коши

На рис. 5 приведена структурная схема настраиваемой модели, обучаемой с помощью процедуры (23).

Эта схема несколько проще той, что приведена на рис. 3, и обе они существенно проще нейросетевой системы робастной идентификации, описанной в [16] и содержащей нелинейные преобразователи и есть расширением на робастный случай алгоритма обучения, введенного в [17].

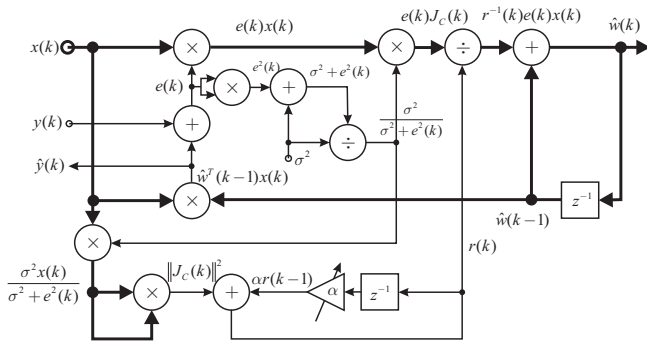


Рис. 5. Настраиваемая модель, обучаемая на основе функции потерь Коши

### Ансамбль гибридных адаптивных моделей идентификации нестационарных систем

С целью исключить эмпирический выбор той или иной модели идентификации нестационарного нелинейного сигнала, загрязненного выбросами с неизвестными законами распределения, целесообразно построить ансамбль адаптивных моделей, который бы обучался в процессе обработки потока данных.

Таким образом, входной сигнал  $X(k) = (x_1(k), x_2(k), \dots, x_n(k))^T$  параллельно обрабатывается  $P$  адаптивными моделями идентификации, формирующими на своих выходах скалярные сигналы  $y_i(k)$ , которые могут быть объединены в  $(P \times 1)$  – вектор выходов  $Y(k) = (y_1(k), y_2(k), \dots, y_p(k))^T$ . Далее этот сигнал поступает на вход оптимизационного блока, где преобразуется в наилучший в смысле принятого внешнего критерия скалярный сигнал  $\bar{y}(k)$ . На рис. 6 представлена архитектура ансамбля адаптивных моделей идентификации.

В большинстве случаев для определения качества решения задач идентификации и прогнозирования принято использовать критерий вида

$$MAPE_i = \frac{100}{N} \sum_{k=1}^N \left| \frac{y(k) - \hat{y}_i(k)}{y(k)} \right|, \quad (24)$$

где  $N$  – количество наблюдений в выборке,  $y$  – фактическое значение сигнала,  $\hat{y}_i$  – значение, полученное на выходе  $i$ -модели,  $MAPE_i$  (*Mean*

*Absolute Percentage Error*) – средняя абсолютная процентная ошибка  $i$ -й модели.

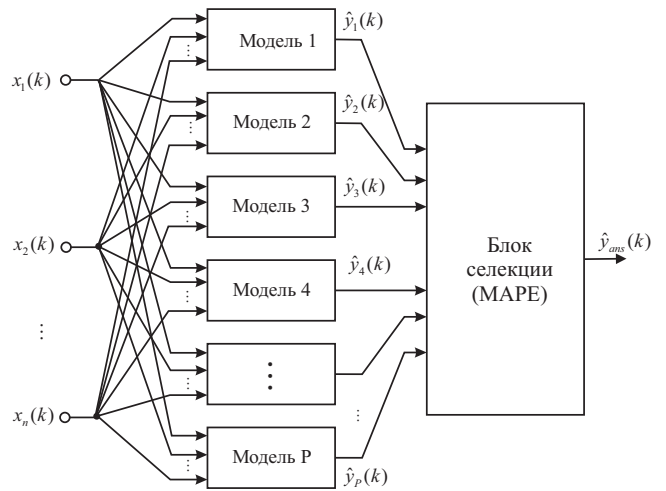


Рис. 6. Ансамбль адаптивных моделей идентификации

Следовательно, в блоке селекции ансамбля адаптивных гибридных моделей в каждый момент дискретного времени будет проводиться выбор наилучшего выходного сигнала  $\hat{y}_i$  в рамках принятого критерия (21) таким образом, что

$$MAPE(k) = \min_i [MAPE_i(k)], \quad (25)$$

$$\hat{y}_{ans}(k) = y_i(k).$$

### Экспериментальное моделирование

Численное моделирование разработанного ансамбля гибридных адаптивных моделей идентификации проводилось на основе сигнала, зашумленного интенсивными выбросами. Сигналы получены на основе ряда Мэки–Гласса, сгенерированными с помощью разностного уравнения

$$\dot{x} = \frac{0,2x(t-\tau)}{1+x^{10}(t-\tau)} - 0,1x(t) \quad (26)$$

(значения временного ряда для каждой целой точки получены с помощью метода Рунге–Кутты четвертого порядка, используемый временной шаг для данного метода выбран 0,1, а начальные условия  $x(0) = 1,2$ , задержка  $\tau = 17$  и  $x(t)$  получены для  $t = 0 \dots 51000$ ) и наложенного на них случайного шума, сгенерированного по распределению Коши.

На рис. 7 представлен общий вид обрабатываемого сигнала с зашумлением по распределе-

нию Коши. Значения  $x(t-18)$ ,  $x(t-12)$ ,  $x(t-6)$ ,  $x(t)$  использованы для идентификации  $x(t+6)$ . Начальные значения синаптических весов сгенерированы случайным образом в диапазоне от  $-0,1$  до  $0,1$ .

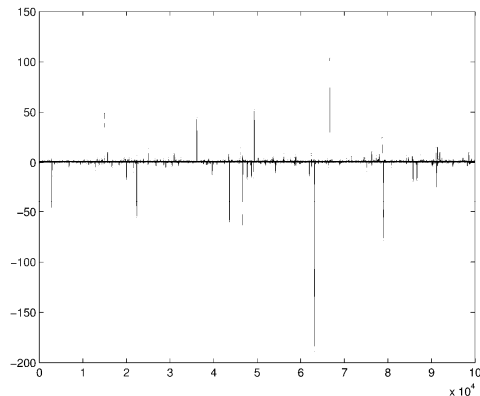


Рис. 7. Временной ряд Мэки–Гласса с шумом, имеющим распределение Коши

В качестве критерия качества использована среднеквадратичная ошибка (*RMSE*).

На рис. 8, *а* представлены результаты идентификации зашумленного временного ряда (реальные значения (пунктирная линия) и значения идентификации (сплошная линия)). На рис. 8, *б* представлен сегмент процесса обучения; как видно, ряд выбросов с большой амплитудой, присутствующих в начале выборки, не оказали большого влияния на метод обучения.

Сравнение результатов идентификации на основе робастного метода обучения проводилось с результатами идентификации на основе методов стохастической аппроксимации и алгоритма на основе рекуррентного метода наименьших квадратов, где структура модели и количество параметров настройки было одинаковым. На рис. 9, *а* представлены результаты идентификации сигнала на основе метода стохастической аппроксимации (фактические значения (пунктирная линия) и значения идентификации (сплошная линия)). На рис. 9, *б* представлен сегмент процесса обучения; как видно, первый же выброс в начале выборки сильно повлиял на процесс обучения и как результат – большая ошибка идентификации.

При обучении модели рекуррентным методом наименьших квадратов при первом выбро-

се происходит *взрыв параметров* ковариационной матрицы и как результат – невозможность идентификации сигналов, зашумленных аномальными выбросами.

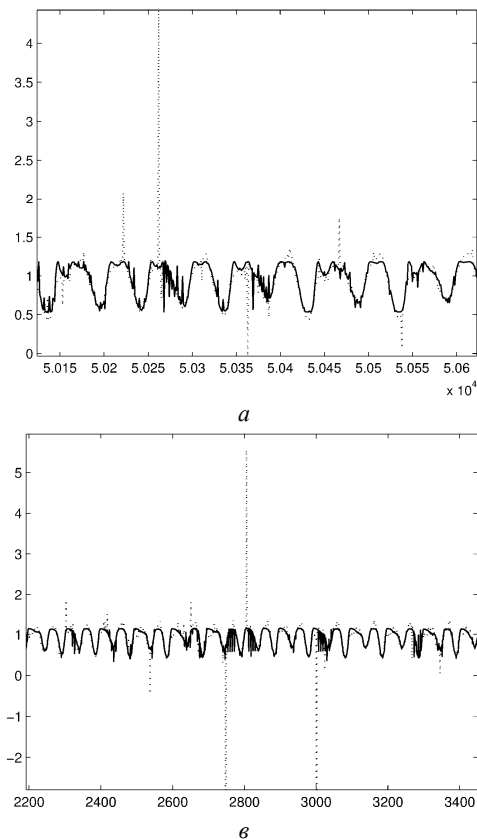


Рис. 8. Результаты обработки зашумленного сигнала с помощью робастного метода обучения

В таблице приведены результаты сравнения идентификации сигнала на основе различных подходов.

Модель и метод обучения	<i>RMSE</i>
Настраиваемая модель, обучаемая с помощью алгоритма Гудвина–Ремеджа–Кэйнеса	0,3203
Настраиваемая модель, обучаемая на основе функции потерь <i>Geman–McClure</i>	0,1932
Настраиваемая модель, обучаемая на основе функции потерь Коши	0,0723
Настраиваемая модель, обучаемая на основе настраивания с помощью рекуррентного метода наименьших квадратов	$\infty$
Ансамбль гибридных адаптивных моделей идентификации	0,0735

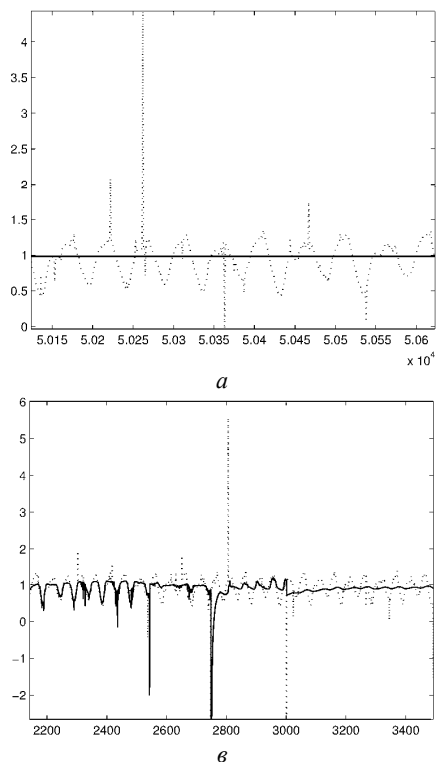


Рис. 9. Результаты обработки зашумленного сигнала с помощью алгоритмов стохастической аппроксимации

**Заключение.** Предложены численно устойчивые и быстродействующие адаптивные алгоритмы робастной идентификации, отличительной особенностью которых является возможность их реализации на основе обучаемых моделей, в основе которых лежат элементарные арифметические операции. Возможность *on-line* обработки поступающей информации обеспечивает решение широкого класса задач, возникающих в рамках интеллектуального анализа потоков данных. Предложен ансамбль гибридных адаптивных моделей идентификации нестационарных систем, позволяющий выбрать наилучшую (в смысле принятого критерия качества) ту или иную модель идентификации сигнала в зависимости от свойств обрабатываемого сигнала. Результаты проведенных экспериментов подтверждают целесообразность использования предложенного ансамбля гибридных адаптивных моделей для решения задач *Data Stream Mining*.

1. Aggarwal C.C., Reddy C.K. Data Clustering // Algorithms and Applications. – Boca Raton: CRC Press, 2014. – 620 p.
2. Aggarwal C.C. Data Mining. – Cham: Springer, 2015. – 734 p.
3. Bifet A. Adaptive Stream Mining: Pattern Learning and Mining from Evolving Data Streams. – IOS Press., 2010. – 224 p.
4. Lughofer E. Evolving Fuzzy Systems. – Berlin-Heidelberg: Springer-Verlag, 2011. – 454 p.
5. Цыпкин Я.З. Основы информационной теории идентификации. – М.: Наука, 1984. – 320 с.
6. Ljung L. System Identification: Theory for the User. – Prentice Hall, 1999. – 672 p.
7. Goodwin Y.C., Ramadge P.J., Caines P.E. A globally convergent adaptive predictor // Automatica. – 1981. – N 17. – P. 135–140.
8. Rey J.W.W. Robust Statistic Methods. – Berlin-Heidelberg–New York: Springer, 1978. – 128 p.
9. Polyak B.T., Tsyypkin Y.Z. Robust identification // Automatica. – 1980. – N 16. – P. 53–63.
10. Хьюбер П. Робастность в статистике. – М.: Мир, 1984. – 304 с.
11. Li S.Z. Markov Random Field Modeling in Computer Vision. – Berlin: Springer, 1995. – 285 p.
12. Zhang Zh. Parameter estimation techniques: a tutorial with application to conic fitting // Image and Vision Comp. J. – 1997. – 15, is. 1. – P. 59–76.
13. Noisy time series production using M-estimate based on robust radial basic function neural networks with growing and pruning techniques / Ch-Ch. Lee, Yu.-Ch. Chiang, Ch-Yu. Shin et al. // Expert Systems with Applications. – 2009. – N 36. – P. 4717–4724.
14. Robust Statistics. The Approach Based on Influence Functions / F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw et al. // Wiley, 2005. – 536 p.
15. Bodyanskiy Ye., Kolodyazhniy V., Stephan A. An adaptive learning algorithm for neural-fuzzy network / Ed. by B. Reusch Computational Intelligence. Theory and Applications. – Berlin–Heidelberg–New York: Springer, 2001. – P. 68–75.
16. Dave R.N., Krishnapuram R. Robust clustering methods: a unified view // IEEE Transaction on Fuzzy Systems. – 1997. – 5. – P. 270–293.
17. Otto P., Bodyanskiy Ye., Kolodyazhniy V. A new learning algorithm for a forecasting neuro-fuzzy network // Integrated Computer – Aided Engineering. – 2003. – 10, № 4. – P. 399–409.

Поступила 20.09.2016

E-mail: bodya@kture.kharkov.ua

© Е.В. Бодянский, Е.А. Винокурова, И.О. Кобылин,  
П.П. Мулеса, 2016

## Robust Adaptive Identification of Non-Stationary Time Series Using Ensemble of Tuning Hybrid Adaptive Models

**Keywords:** adaptive identification, robust function, non-linear non-stationary time series, ensemble of hybrid models, outliers with unknown distribution law.

**Introduction.** A synthesis of monitoring and control systems for stochastic plants under uncertain conditions requires an identification circuit in these systems, which provides the continuous tuning of mathematical models based on observational data for input and output signals in real time.

**Purpose.** For solving this problem an effective tools are methods of the informational theory of identification, forming a common approach, which is connected with recurrent algorithms of optimization under uncertain conditions.

These algorithms tune mathematical models based on input and output signals of system. Different types of such systems (both static and dynamic) can be represent as a pseudo-linear regression equation  $y(k) = w^T x(k) + \zeta(k)$  where  $y(k)$  is scalar output of object in discrete time  $k = 1, 2, \dots, N$ ,  $w = (w_0, w_1, \dots, w_n)^T$  is  $(n+1) \times 1$  vector of unknown parameters (regression coefficients), which should be estimated,  $x(k) \in R^n$  is a vector of input variables,  $\zeta(k)$  is random noise with unknown function of the density distribution.

Most of procedures and algorithms for the identification usually are based on the hypothesis of Gaussian distribution of noise  $\zeta(k) \sim N(0, \sigma^2)$ . It led to using of the least squares method in the different modifications. If the data are fed sequentially for processing or the sample size  $N$  is not fixed, preference should be given to the recurrent least squares method, which, however, for the large values of  $n$  can be numerically unstable, tedious and ultimately leads to «blow-up» of the covariance matrix of the parameters, because a computational procedure of matrix inverse may be unstable.

**Conclusions.** Therefore, computationally simple and high-speed adaptive algorithms for robust identification of non-stationary non-linear time series are proposed. The distinctive feature of such algorithms is an ability to implement them using the learning models, which are consisted of the elementary arithmetic operations. The possibility of on-line information processing provides a solution of a wide class of problems, which are appeared in Data Stream Mining tasks. Also an ensemble of tuning hybrid adaptive models is proposed. This system allows to choose the best model in the context of accepted quality criterion at each discrete instant of time.



**Внимание !**

**Оформление подписки для желающих  
опубликовать статьи в нашем журнале обязательно.**

**В розничную продажу журнал не поступает.**

**Подписной индекс 71008**