

УДК 62.505: 621.391.8

О МЕТОДАХ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ И НЕПРЕРЫВНЫХ ПРОЦЕССОВ**Ермолаев Валерий Андреевич**

кандидат технических наук, доцент кафедры электроники и вычислительной техники Муромского института (филиала) ФГБОУ ВО «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых».

E-mail: valeermolaev@yandex.ru.

Адрес: 602264, г. Муром, ул. Орловская, 23.

Аннотация: Установлены возможности алгоритмов прогнозирования различных классов. Показаны ограниченные возможности методов декомпозиции процессов на эмпирические моды и методы параметрического прогнозирования, основанные на представлении временного ряда обобщённым многочленом по системе линейно независимых функций. Получены алгоритмы прогноза, основанные на моделях авторегрессии. Получены рекуррентные уравнения, определяющие явную зависимость оценок прогноза от коэффициентов модели. Поставлена задача нахождения оценок прогноза посредством минимизации функции потерь – квадрата нормы отклонения оценок от наблюдаемых значений временного ряда. Получен способ обобщения алгоритма прогноза с помощью линейной модели авторегрессии на нелинейные модели – модели, представленные функциональными рядами или системами Вольтерра.

Ключевые слова: прогнозирование, горизонт предсказания, авторегрессия, функциональный ряд, нелинейная система Вольтерра, нейронная сеть.

Введение

Задача прогнозирования или предсказания заключается в нахождении будущих значений временного ряда или непрерывного процесса на интервале, называемом горизонтом предсказания [1, 2], в пределах которого обеспечивается необходимая точность решения задачи. Для непрерывных процессов – это интервал $(t, t + \tau]$, который для временного ряда, полученного, в частности, посредством дискретизации непрерывного процесса, записывается как $(n, n + N]$. Здесь t и nT_0 – текущие моменты времени, при этом T_0 – период дискретизации. Прогнозирование обычно осуществляется по значениям временного ряда или процесса на конечном, предшествующем, интервале $[t - T, t]$ времени.

Горизонт предсказания является не только одной из важнейших мер качества предсказуемости, но и используется в качестве критерия степени детерминированности и случайности наблюдаемых явлений, служит характеристикой динамического хаоса (характеристикой хаотических колебаний в динамических системах) [1, 2]. В основе этого утверждения лежит зависимость горизонта предсказания не только

от используемых алгоритмов, но и от свойств анализируемых временных рядов и процессов. В задачах собственно прогнозирования выбор алгоритма осуществляется исходя из соображений максимизации горизонта предсказания.

Один из принципов прогнозирования временных рядов или процессов основывается на их представлении непрерывными или дискретными моделями. В работах [1, 2] модель схематически описывается дифференциальным уравнением, зависящим от неизвестных параметров системы a и факторов f_k , отражающих неопределённость модели. При этом в рассмотрение вводятся три процесса: наблюдаемый процесс y , исследуемый (истинный) процесс x и модельный (прогностический) процесс z .

Истинный процесс в силу неопределённости факторов f_k является (из множества возможных) неизвестным решением дифференциального уравнения

$$P(d/dt, x, a, f_k) = 0.$$

Пренебрежение указанными факторами позволяет получить дифференциальное уравнение, описывающее модельный процесс z ,

при соответствующих условиях близкий к истинному процессу x . Это уравнение можно записать в виде

$$G(d/dt, z, a) = 0.$$

Простейшим критерием качества прогнозирования является, при этом, среднеквадратическое по факторам f_k значение нормы отклонения модельного процесса от истинного на интервале предсказания, то есть величина

$$\langle \|x - z\|^2 \rangle = \frac{1}{M} \sum_{k=1}^M \int_t^{t+\tau} |x(v, f_k) - z(v)|^2 dv.$$

Этот и другие критерии рассматриваются в [1]. Конечно, здесь существует проблема соответствия истинного и наблюдаемого процессов, то есть проблема построения модели по наблюдаемым данным и решения множества дифференциальных уравнений, соответствующих различным факторам f_k .

Аналогично определяется качество предсказуемости дискретных процессов или временных рядов, представленных дискретными моделями, которые, в частности, могут быть получены из непрерывных моделей путём замены производных конечными разностями. При этом операция интегрирования заменяется операцией суммирования по конечному на интервале прогнозирования множеству данных. Переход к дискретной модели эквивалентен численному решению дифференциального уравнения со свойственными этому решению проблемами чувствительности к возмущающим воздействиям.

На практике построение моделей истинных процессов основывается, так или иначе, на данных о соответствующих наблюдаемых процессах: модели могут относиться к классам линейных дискретных [3] и регрессионных [4] систем. Использование подобных моделей при прогнозировании осложнено целым рядом факторов, обусловленных наличием выбросов, негауссовским характером распределения ошибок наблюдения и их коррелированностью, зависимостью параметров моделей от

времени и неадекватностью моделей наблюдаемым данным [4].

Сложность прогнозирования нестационарных процессов обусловила появление метода, сочетающего в себе их декомпозицию на эмпирические моды (empirical mode decomposition – EMD) [5 – 8] и преобразование Гильберта [5, 9]. Этот метод обеспечивает приближение наблюдаемого процесса суммой слагаемых вида $A_k(t)\exp(j\varphi_k(t))$, прогнозирование каждого из которых становится более надёжным и простым. В совокупности операция EMD и преобразование Гильберта называется преобразованием Гильберта – Хуанга [5, 6]. При этом за эквивалент операции EMD принимаются и другие средства декомпозиции, в частности, банки фильтров [10].

Распространёнными методами собственно прогнозирования являются параметрические методы регрессионной аппроксимации [4, 11, 12], динамические модели авторегрессии – скользящего среднего или скользящего среднего [3, 12 – 15], методы импульсных функций [16], нелинейные модели Вольтерра [17, 18] и искусственные нейронные сети [19 – 22]. Решение задачи прогнозирования осуществляется не только в дискретном, но и в непрерывном времени [23].

Следует отметить, что в рамках каждого из указанных подходов существует множество различных возможностей или вариантов решения задачи прогнозирования. Настоящая работа в этом плане преследует цель охарактеризовать ряд из известных вариантов по их достоинствам и недостаткам, а также представить новый алгоритм предсказания, аналогичный алгоритму локальной аппроксимации на искусственных нейронных сетях [24].

1. Параметрическое регрессионное прогнозирование

Прогноз временного ряда этим методом предполагает нахождение функции регрессии или обобщённого многочлена по системе линейно независимых функций $\{\varphi_k\}_1^n$. Многочлен нахо-

дится методом наименьших квадратов по известным на текущем интервале данным. Многочлен может быть записан в виде

$$P(t) = \sum_{k=1}^n a_k \varphi_k(t) = \Phi^T(t) \mathbf{a}, \quad (1)$$

где $\Phi = (\varphi_1, \varphi_2, \dots, \varphi_n)^T$ и $\mathbf{a} = (a_1, a_2, \dots, a_n)^T$.

Если ввести матрицу, столбцами которой являются векторы значений базисных функций в узлах данных, то есть матрицу

$$\Phi = (\Phi(t_1), \Phi(t_2), \dots, \Phi(t_N)),$$

то вектор коэффициентов искомого многочлена

$$\mathbf{a} = (\Phi \Phi^T)^{-1} \Phi \mathbf{y}, \quad (2)$$

где $\mathbf{y} = (y_1, y_2, \dots, y_N)^T$ – вектор наблюдаемых в текущий момент данных.

При этом за искомым прогнозом временного ряда в момент $t_N + \tau$ берётся [4] величина

$$\hat{y}(t_N + \tau) = \Phi^T(t_N + \tau) \mathbf{a}.$$

К сожалению, горизонт предсказания τ полученной таким образом оценки невелик, чтобы её использовать на практике. Кроме того, эта оценка весьма чувствительна к степени многочлена n , что выражается в значительном росте ошибки прогнозирования при увеличении степени n . По крайней мере, эта закономерность имеет место для многочленов по степенной системе базисных функций, так что $\Phi = (1, t, t^2, \dots, t^{n-1})^T$.

Параметрический метод можно, как представляется, применить, если на интервале прогнозирования функциональный ряд по системе базисных функций $\{\varphi_k\}_1^\infty$ является сходящимся, а всеми его слагаемыми, начиная с номера $n+1$, можно пренебречь. Кроме того, функции системы $\{\varphi_k\}_1^n$ должны, желательно, иметь на интервале их определения примерно одинаковую вариацию. В этом плане интересна система функций, ортонормированных на интервале $[0, \infty)$. Следует, однако, отметить, что приведённые предположения, требуют ещё своего доказательства.

Данный метод является по существу методом статической локальной аппроксимации, в котором никак не отражается динамика временного ряда. Учесть эту динамику в принципе можно, установив зависимость коэффициентов a_k обобщённого многочлена (1) от времени, точнее, от номера окна локальной аппроксимации. Этим, фактически, осуществляется декомпозиция временного ряда на множество переменных коэффициентов многочлена (1), что позволяет отнести эту операцию к классу операций EMD.

Горизонт предсказуемости можно увеличить, если дополнить функцию потерь, минимизация которой даёт формулу (2), регуляризирующим функционалом, стабилизирующим решение на интервале прогнозирования. При этом коэффициенты многочлена можно, например, получить, решая задачу минимизации функции

$$\sum_{l=1}^N (\Phi^T(t_l) \mathbf{a} - y_l)^2 + \alpha \int_{t_N - \nu}^{t_N + \tau} (q \cdot \Phi^T(t) \mathbf{a})^2 dt,$$

где q – некоторый, в частности дифференциальный, оператор. Как и выше, здесь не учитывается динамический характер временного ряда, а проблема выбора коэффициента регуляризации α и интервала интегрирования $[t_N - \nu, t_N + \tau]$ остается открытой.

2. Динамические модели прогнозирования

Задача прогнозирования при определённых условиях сводится к построению динамической модели, например, авторегрессии или авторегрессии–скользящего среднего, то есть к задаче представления временного ряда процессом на выходе некоторой динамической системы. Теоретические и чисто математические аспекты указанного представления, а точнее–представления временных рядов линейными системами, освещаются в [25]. В этой работе, в частности, освещаются проблемы точного и приближённого моделирования, а также рассматриваются модели многомерной авторегрессии и модели в переменных состояниях.

Прогноз в рамках подобного подхода осуществляется по известному состоянию модели в текущий момент времени и известному входному воздействию.

Линейная дискретная модель, задаваемая импульсной функцией, описывается [16] уравнением

$$x(n) = \sum_{l=-\infty}^n h(n-l)u(l).$$

Здесь $h(k)$ и $u(k)$ – соответственно импульсная функция и входное воздействие модели.

Усечение правой части приведённого уравнения до конечного числа слагаемых даёт [3] модель скользящего среднего:

$$x(n) = \sum_{l=n-m}^n h(n-l)u(l) \equiv \sum_{l=0}^m b_l u(n-l).$$

Модель авторегрессии–скользящего среднего описывается разностным уравнением [3]

$$x(n) = \sum_{l=1}^p \alpha_l x(n-l) + \sum_{l=0}^m b_l u(n-l). \quad (3)$$

Частной формой уравнения (3) является уравнение авторегрессии [3]

$$x(n) = \sum_{l=1}^p \alpha_l x(n-l) + u(n). \quad (4)$$

Если входное воздействие известно, то уравнение (3), включая и его частный случай – уравнение (4), это фактически уравнения прогноза на один шаг вперед. Коэффициенты уравнений (3) и (4) находятся посредством идентификации системы на текущем интервале данных [12, 26, 27].

Общее решение уравнения (3) или (4) с постоянными коэффициентами является, как известно [3, 28], суммой его частного решения $f(n)$ и общего решения однородного уравнения. Если линейно независимые частные решения однородного уравнения обозначить как $\varphi_l(n)$, $l=1, 2, \dots, r$, то его общее решение можно записать в виде

$$\varphi(n) = \sum_{k=1}^p c_k \varphi_k(n).$$

Функции $\varphi_k(n)$ выражаются через собственные значения характеристического многочлена

$$q^n - \sum_{l=1}^p \alpha_l q^{r-l} = 0.$$

При этом, если корень λ_k этого многочлена простой, то собственная функция

$$\varphi_k(n) = \lambda_k^n \equiv \rho_k^n e^{s_k n}.$$

Корню же кратности r_k отвечают функции

$$\varphi_{k,1}(n) = \lambda_k^n \equiv \rho_k^n e^{s_k n},$$

$$\varphi_{k,2}(n) = n \lambda_k^n \equiv n \rho_k^n e^{s_k n},$$

...

$$\varphi_{k,r_k}(n) = n^{r_k-1} \lambda_k^n \equiv n^{r_k-1} \rho_k^n e^{s_k n}.$$

Частное решение неоднородного уравнения зависит от вида сигнала $u(n)$ (см. [28]).

В качестве модели прогноза на один шаг вперед естественно выбрать автономную систему, описываемую уравнением

$$\hat{x}(n) = \sum_{l=1}^p \alpha_l x(n-l). \quad (5)$$

Здесь $\hat{x}(n)$ – прогноз на один шаг. Оценки коэффициентов модели α_k можно, например, получить, используя соответствующие (5) уравнения Юла – Уокера [3, 26].

Если принять, что оценки $\hat{\alpha}_k$ являются действительными значениями коэффициентов модели, то решение задачи прогноза на любое число шагов вперед может быть получено рекуррентным способом по формуле

$$\hat{x}(n+l) = \sum_{j=1}^l \hat{\alpha}_j \hat{x}(n-j) + \sum_{j=l+1}^p \hat{\alpha}_j x(n-j), \quad (6)$$

$$l = 0, 1, \dots$$

К сожалению, оценки прогноза (6) и коэффициенты модели $\hat{\alpha}_k$, полученные в ходе прогнозирования на один шаг, не являются оптимальными в случае прогноза на большее число шагов, $k \geq 1$, когда более естественной мерой погрешности является величина нормы

$$e_n = \|\mathbf{x}_n - \hat{\mathbf{x}}_n\|,$$

где $\mathbf{x}_n = (x_n, x_{n+1}, \dots, x_{n+k})^T$, $x_{n+l} \equiv x(n+l)$,

и $\hat{\mathbf{x}}_n = (\hat{x}_n, \hat{x}_{n+1}, \dots, \hat{x}_{n+k})^T$, $\hat{x}_{n+l} \equiv \hat{x}(n+l)$.

Пусть задача прогноза заключается в нахождении $k+1$ оценок $\hat{x}(n+l)$ при значениях $l = 0, 1, \dots, k$.

В соответствии с уравнением (6) оценки $\hat{x}(n+l)$ находятся последовательно, начиная со значения $l=0$ и кончая $l=k$.

Если ввести вектор коэффициентов модели $\mathbf{a} = (\alpha_0, \alpha_1, \dots, \alpha_r)^T$, матрицу

$$\hat{\mathbf{X}}_n = \begin{pmatrix} x_{n-1} & x_{n-2} & x_{n-3} & \dots & x_{n-p} \\ \hat{x}_n & x_{n-1} & x_{n-2} & \dots & x_{n-p+1} \\ \hat{x}_{n+1} & \hat{x}_n & x_{n-1} & \dots & x_{n-p+2} \\ \vdots & \vdots & \vdots & & \vdots \\ \hat{x}_{n+p-2} & \hat{x}_{n+p-3} & \hat{x}_{n+p-4} & \dots & x_{n-1} \\ \hat{x}_{n+p-1} & \hat{x}_{n+p-2} & \hat{x}_{n+p-3} & \dots & \hat{x}_n \\ \vdots & \vdots & \vdots & & \vdots \\ \hat{x}_{n+k-1} & \hat{x}_{n+k-2} & \hat{x}_{n+k-3} & \dots & \hat{x}_{n+k-p} \end{pmatrix}$$

и вектор прогноза $\hat{\mathbf{x}}_n = (\hat{x}_n, \hat{x}_{n+1}, \dots, \hat{x}_{n+k})^T$, то уравнение (6) на интервале $[n, n+k]$ может быть записано в виде

$$\hat{\mathbf{x}}_n = \hat{\mathbf{X}}_n \mathbf{a}. \tag{7}$$

Если ввести вектор $\mathbf{y}_n = (y_n, y_{n+1}, \dots, y_{n+k})^T$ данных, наблюдаемых на интервале $[n, n+k]$ прогнозирования, и матрицу \mathbf{Y}_n , полученную из матрицы $\hat{\mathbf{X}}_n$ посредством замены её компонентов \hat{x}_{n+l} и x_{n+l} на наблюдаемые данные y_{n+l} , то коэффициенты модели в соответствии с (7) можно найти из уравнения

$$\mathbf{y}_n = \mathbf{Y}_n \mathbf{a}. \tag{8}$$

Исходя из уравнения (7) коэффициенты модели можно найти и иным способом, а именно, посредством минимизации функционала

$$J(\mathbf{a}) = \|\mathbf{y}_n - \hat{\mathbf{x}}_n\|^2 = \|\mathbf{y}_n - \hat{\mathbf{X}}_n \mathbf{a}\|^2, \tag{9}$$

где компоненты матрицы $\hat{\mathbf{X}}_n$, отвечающие прогнозируемым значениям \hat{x}_{n+l} временного ряда, должны быть представлены в форме явных функций от коэффициентов модели.

Учитывая, что второй и последующие компоненты вектора, определяемого произведением $\hat{\mathbf{X}}_n \mathbf{a}$, являются нелинейными функциями коэффициентов модели, минимизация функции (9) требует применения численных методов.

Явный вид зависимости оценок прогноза от коэффициентов модели можно получить посредством рекуррентного решения уравнения (7).

Введём переменные

$$\sigma_l = \sum_{j=l+1}^p \alpha_j x_{n-j+1}, \quad l = 0, 1, \dots, p-1.$$

Нетрудно показать, что оценки прогноза являются линейными комбинациями этих переменных, то есть

$$\hat{x}_{n+l} = \sum_{j=0}^l d_{l-j+1} \sigma_j, \quad l = 0, 1, \dots. \tag{10}$$

Коэффициенты уравнения (10) находятся при этом рекуррентным способом:

$$d_l = \sum_{j=1}^l \alpha_j d_{l-j}, \quad d_0 = 0, \quad d_1 = 1.$$

Так, например, отсюда получаем

$$\begin{aligned} d_2 &= \alpha_1, \\ d_3 &= \alpha_1^2 + \alpha_2, \\ d_4 &= \alpha_1^3 + 2\alpha_1\alpha_2 + \alpha_3, \\ d_5 &= \alpha_1^4 + 3\alpha_1^2\alpha_2 + 2\alpha_1\alpha_3 + \alpha_2^2 + \alpha_4, \dots \end{aligned}$$

Здесь следует отметить, что если величина $l \geq p$, то переменная $\sigma_l = 0$ и оценка прогноза принимает вид

$$\hat{x}_{n+l} = \sum_{j=0}^p d_{l-j+1} \sigma_j.$$

Аналогичные, только более громоздкие, выражения могут быть получены и для модели авторегрессии – скользящего среднего.

Оценивание коэффициентов модели временного ряда, как с помощью уравнения (8), так и посредством минимизации функционала (9), предполагает, что наблюдаемые данные порождаются детерминированным процессом. В стохастическом случае возможно применение техники корреляционного анализа, анало-

гичной технике, используемой при выводе уравнений Юла – Уокера [3, 26].

Конечно, с помощью линейных моделей невозможно описать все возможные временные ряды, к числу которых относятся, в частности, ряды, порождаемые нелинейными системами. В подобных случаях естественным представляется применение нелинейных моделей, если они обеспечивают большее соответствие действительности и более надёжные результаты прогнозирования. Наибольшее распространение в качестве таковых сегодня получили модели типа систем Вольтерра и нейронных сетей, число параметров которых, следует отметить, заведомо превышает размерности входного и выходного векторов данных.

3. Нелинейные модели прогноза.

Модели Вольтерра

Область применения нелинейных систем или, по-другому, функциональных рядов Вольтерра чрезвычайно обширна. Поток работ по этой теме со временем только расширяется. Здесь отметим монографию [29] и несколько работ [17, 18, 30]. Известно также множество монографических изданий и журнальных публикаций, в основном на английском языке, по теории и применению систем Вольтерра в задачах анализа и синтеза электронных устройств, а также в задачах нелинейной адаптации и компенсации помех и искажений.

Понятно, что в рамках указанного подхода дискретную модель наблюдаемых данных следует представить конечным числом слагаемых функционального ряда

$$y(n) = d_0 + \sum_{r_1=0}^p h_{r_1} x(n-r_1) + \sum_{r_1=0}^p \sum_{r_2=0}^p h_{r_1, r_2} x(n-r_1)x(n-r_2) + \dots + \sum_{r_1=0}^p \dots \sum_{r_k=0}^p h_{r_1, \dots, r_k} x(n-r_1) \dots x(n-r_k).$$

Если ввести вектор выходов системы $\mathbf{y} = (y(0) \ y(1) \ \dots \ y(N))^T$,

то, совместно упорядочив коэффициенты модели h_{r_1, \dots, r_l} и соответствующие им множители $x(n-r_1) \dots x(n-r_l) \equiv x_{n-r_1} \dots x_{n-r_l}$, можно получить уравнение [18]

$$\mathbf{y} = \mathbf{X}^T \mathbf{h}.$$

Здесь [18] вектор коэффициентов

$$\mathbf{h} = (d_0, h_0, \dots, h_p, h_{0,0}, \dots, h_{p,p}, \dots, h_{p, \dots, p})^T$$

размера $M \times 1$ и матрица размера $M \times N$

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & \dots \\ x_0 & x_1 & x_2 & \dots \\ x_{-1} & x_0 & x_1 & \dots \\ \vdots & \vdots & \vdots & \ddots \\ x_{n-p} & x_{n-p+1} & x_{n-p+2} & \dots \\ x_0^2 & x_1^2 & x_2^2 & \dots \\ x_0 x_{-1} & x_1 x_0 & x_2 x_1 & \dots \\ \vdots & \vdots & \vdots & \ddots \\ x_{n-p}^k & x_{n-p+1}^k & x_{n-p+2}^k & \dots \end{pmatrix}.$$

Если принять, что $N = M$, то коэффициенты модели можно найти из уравнения

$$\hat{\mathbf{h}} = (\mathbf{X}^T)^{-1} \mathbf{y}.$$

Соответственно и оценки временного ряда определяются выражением

$$\hat{\mathbf{x}} = \mathbf{X}^T \hat{\mathbf{h}}. \quad (11)$$

В случае когда $N > p$, компоненты \hat{x}_n вектора оценок (11), начиная с номера $n > p$, являются оценками прогноза на несколько, а именно $n - p + 1$, шагов вперед.

На практике обычно используются нелинейные модели второго и третьего порядка ($k = 2$ или 3) в силу резкой зависимости сложности задачи от порядка модели, а именно от числа её коэффициентов

$$K = \sum_{l=1}^k C_{p+l}^l,$$

где C_N^m – общепринятое обозначение числа сочетаний из N элементов по m .

В принципе, на модели Вольтерра можно распространить и результаты предыдущего раздела, в частности, метод нахождения коэффициентов линейной модели (9), позволяющий

уменьшить объём данных, необходимых для их оценивания.

4. Нейросетевые методы прогнозирования

Горизонт предсказания любой модели прогноза зависит от того, насколько точно эта модель воспроизводит динамику временного ряда или системы, порождающей наблюдаемый процесс. В этой связи возникает, в частности, вопрос о горизонте предсказания моделей в форме искусственных нейронных сетей прямого распространения. Вопрос здесь заключается в том, насколько точно динамика процесса может быть представлена весовыми коэффициентами сети в отсутствии обратных связей. Конечно, в этом случае отсутствие обратных связей в какой-то степени компенсируется алгоритмами обучения типа обратного распространения ошибки.

Алгоритм функционирования многослойной нейронной сети прямого распространения при прохождении сигналов по направлению от входа к выходу зададим уравнениями в матричной форме.

Введем вектор выходов j -слоя сети

$$\mathbf{y}_j = (y_{j,1}, y_{j,2}, \dots, y_{j,m_j})^T$$

и вектор весовых коэффициентов l -нейрона в j -слое сети

$$\mathbf{w}_{j,l} = (w_{j,l,1}, w_{j,l,2}, \dots, w_{j,l,m_{j-1}})^T.$$

Тогда синаптическая сумма l -нейрона в j -слое сети

$$s_{j,l} = \mathbf{w}_{j,l}^T \mathbf{y}_{j-1} + w_{j,l,0},$$

где $w_{j,l,0}$ – смещение нейрона.

При этом выходы j -слоя сети можно представить вектором

$$\mathbf{y}_j = \begin{pmatrix} \varphi(\mathbf{w}_{j,1}^T \mathbf{y}_{j-1} + w_{j,1,0}) \\ \varphi(\mathbf{w}_{j,2}^T \mathbf{y}_{j-1} + w_{j,2,0}) \\ \vdots \\ \varphi(\mathbf{w}_{j,m_{j-1}}^T \mathbf{y}_{j-1} + w_{j,m_{j-1},0}) \end{pmatrix}, \quad (12)$$

где $\varphi(\cdot)$ – функция активации нейронов сети.

Формула (12) является, фактически, рекуррентным уравнением, позволяющим найти последовательно выходы всех слоёв сети, начиная с первого слоя, $j=1$, и кончая последним, $j=n$, слоем, совпадающим с выходом сети. Вектор \mathbf{y}_0 при этом – это известный выход сенсорного слоя сети.

При решении задач прогноза выходами сети в пределах одного цикла её функционирования являются результаты предсказания временного ряда или процесса на заданное число шагов вперёд, начиная с 1 и кончая m_n .

Алгоритм обучения нейронной сети методом обратного распространения ошибки также можно представить уравнениями в матричной форме.

Введём матрицу весов j -слоя сети

$$\mathbf{W}_j = (\mathbf{w}_{j,1}, \mathbf{w}_{j,2}, \dots, \mathbf{w}_{j,m_j}),$$

вектор смещения нейронов j -слоя

$$\mathbf{w}_{j,0} = (w_{j,1,0}, w_{j,2,0}, \dots, w_{j,m_j,0})^T,$$

вектор синаптических сумм

$$\mathbf{s}_j = (s_{j,1}, s_{j,2}, \dots, s_{j,m_j})^T$$

и векторную функцию активации j -слоя

$$\boldsymbol{\varphi}(\mathbf{s}_j) = (\varphi(s_{j,1}), \varphi(s_{j,2}), \dots, \varphi(s_{j,m_j}))^T.$$

Весовые коэффициенты j -слоя сети находятся при этом в ходе минимизации функционала

$$J_j(\mathbf{e}_j) = (\boldsymbol{\varphi}(\mathbf{s}_j) - \bar{\mathbf{y}}_j)^T (\boldsymbol{\varphi}(\mathbf{s}_j) - \bar{\mathbf{y}}_j),$$

где $\bar{\mathbf{y}}_j = (\bar{y}_{j,1}, \bar{y}_{j,2}, \dots, \bar{y}_{j,m_j})^T$ – вектор требуемых выходов j -слоя сети и

$$\boldsymbol{\varphi}(\mathbf{s}_j) - \bar{\mathbf{y}}_j = \mathbf{e}_j -$$

вектор ошибок по выходам j -слоя.

Вывод алгоритма обратного распространения ошибки основывается на допущении, что суммарные значения ошибок на выходах слоёв одинаковы, то есть

$$J_{j-1}(\mathbf{y}_{j-1} - \bar{\mathbf{y}}_{j-1}) = J_j(\boldsymbol{\varphi}(\mathbf{s}_j) - \bar{\mathbf{y}}_j).$$

Дифференцируя это равенство с учётом (12) по компонентам вектора \mathbf{y}_{j-1} получаем

$$2\mathbf{e}_{j-1} = \frac{\partial \mathbf{s}_j^T}{\partial \mathbf{y}_{j-1}} \frac{\partial \Phi^T(\mathbf{s}_j)}{\partial \mathbf{s}_j} \frac{\partial J_j(\Phi(\mathbf{s}_j))}{\partial \Phi^T(\mathbf{s}_j)},$$

где $\partial/\partial \mathbf{z} = (\partial/\partial z_1, \partial/\partial z_2, \dots, \partial/\partial z_n)^T$ – оператор градиента.

Если, далее, ввести матрицы

$$\mathbf{W}_j = \frac{\partial \mathbf{s}_j^T}{\partial \mathbf{y}_{j-1}},$$

$$\Phi_j = \frac{\partial \Phi^T(\mathbf{s}_j)}{\partial \mathbf{s}_j} =$$

$$= \text{diag} \left(\frac{\partial \varphi(s_{j,1})}{\partial s_{j,1}}, \frac{\partial \varphi(s_{j,2})}{\partial s_{j,2}}, \dots, \frac{\partial \varphi(s_{j,m_j})}{\partial s_{j,m_j}} \right),$$

то алгоритм обучения методом обратного распространения ошибки принимает вид разностного уравнения

$$\mathbf{e}_{j-1} = \mathbf{W}_j \Phi_j \mathbf{e}_j, \quad j = n, n-1, \dots, 2, \quad (13)$$

с начальными условиями

$$\mathbf{e}_n = \Phi(\mathbf{s}_n) - \bar{\mathbf{y}}_n.$$

Здесь $\bar{\mathbf{y}}_n = \mathbf{d}_n$ – вектор требуемых результатов прогноза временного ряда, известных на момент обучения сети.

Коррекция весовых коэффициентов сети по ошибкам, полученным с помощью уравнения (13), осуществляется, если воспользоваться градиентным методом, по итерационным формулам

$$\mathbf{w}_{j,l}(k+1) = \mathbf{w}_{j,l}(k) - \alpha \rho_{j,l}(s_{j,l}) e_{j,l} \mathbf{y}_{j-1},$$

$$\mathbf{w}_{j,0}(k+1) = \mathbf{w}_{j,0}(k) - \alpha \Phi_j \mathbf{e}_j, \quad (14)$$

где $l = 1, 2, \dots, m_j$, $e_{j,l}$ является l -компонентом вектора \mathbf{e}_j ошибки, функция

$$\rho_{j,l}(s_{j,l}) = \frac{\partial \varphi(s_{j,l})}{\partial s_{j,l}}$$

и шаг настройки весовых коэффициентов α выбирается из диапазона $0 < \alpha < 1$.

По завершении обучения согласно (12) находится вектор \mathbf{y}_n прогнозируемых значений временного ряда, числом m_n , по вектору \mathbf{y}_0 данных на выходах сенсорного слоя сети. Число m_n в этом случае – это глубина прогно-

за, которую можно увеличить до величины кратной m_n , если воспользоваться процедурой обратной подстановки, в ходе которой осуществляется сдвиг компонентов вектора \mathbf{y}_0 на величину m_n с подстановкой на освободившееся место компонентов вектора прогноза \mathbf{y}_n при неизменных весовых коэффициентах сети.

Пусть размерность сенсорного слоя сети $m_0 \geq m_n$. Введем матрицу размера $m_0 \times m_0$

$$\mathbf{H}_{m_0 \times m_0} = \begin{pmatrix} \mathbf{0}_{m_n \times (m_0 - m_n)} & \mathbf{0}_{m_n \times m_n} \\ \mathbf{I}_{(m_0 - m_n) \times (m_0 - m_n)} & \mathbf{0}_{(m_0 - m_n) \times m_n} \end{pmatrix},$$

где $\mathbf{0}_{i \times i}$ и $\mathbf{I}_{i \times i}$ – соответственно нулевая и единичная матрицы, индексы которых определяют их размерность.

Тогда названной процедуре обратной подстановки соответствует уравнение

$$\mathbf{y}_0(l+1) = \mathbf{H}_{m_0 \times m_0} \mathbf{y}_0(l) + \begin{pmatrix} \mathbf{y}_n(l) \\ \mathbf{0}_{(m_0 - m_n) \times 1} \end{pmatrix}, \quad (15)$$

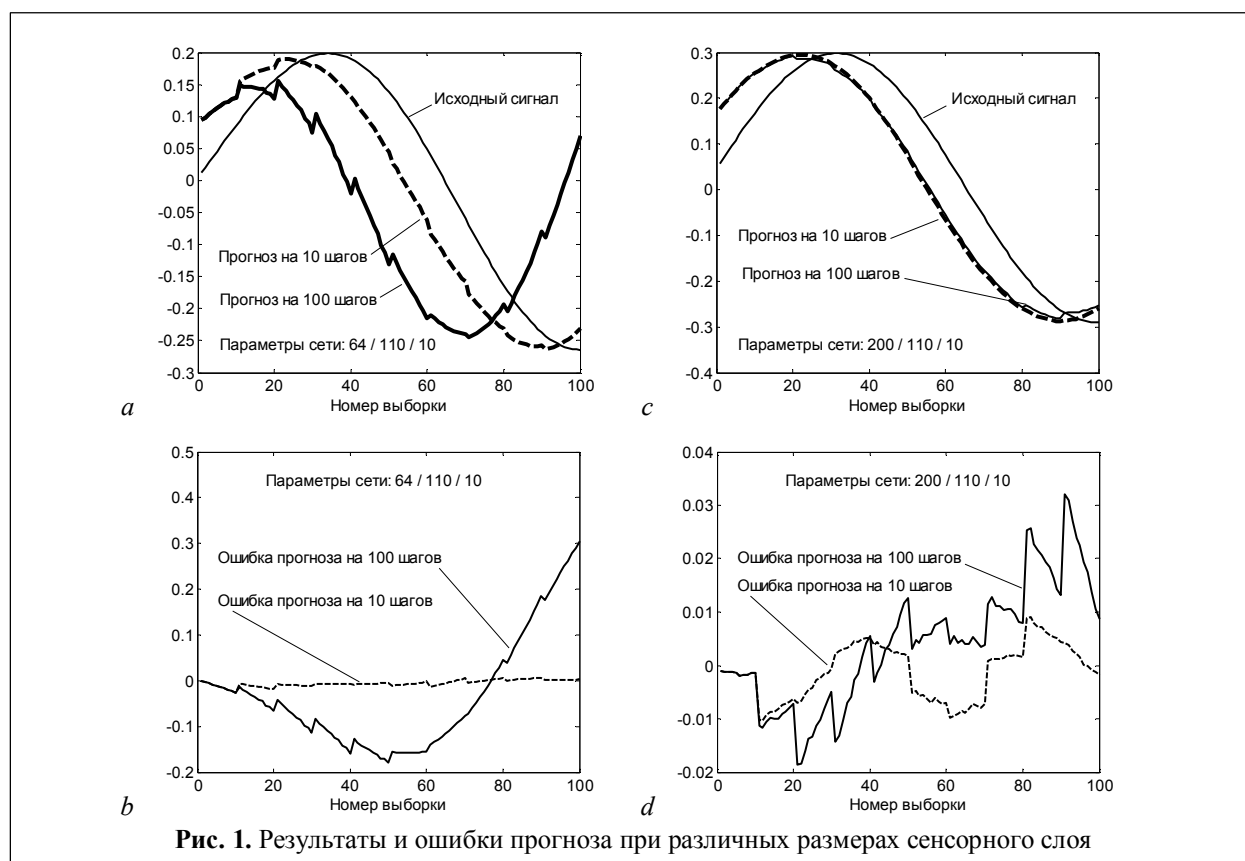
$$l = 1, 2, \dots, k.$$

Здесь $\mathbf{y}_n(l)$ ответ нейронной сети на вектор входов $\mathbf{y}_0(l)$ (ответ, соответствующий глубине прогноза $l \cdot m_n$), при этом $\mathbf{y}_0(1) \equiv \mathbf{y}_0$, а оценка весовых коэффициентов сети производится только в момент, предшествующий нахождению вектора $\mathbf{y}_n(1)$.

Ошибки прогноза, естественно, зависят от того, насколько размеры сенсорного слоя соответствуют характерной величине интервала временного ряда, по которому можно восстановить его динамику, что подтверждается графиками на рис. 1.

При моделировании, результаты которого приведены рис. 1, в качестве функции активации выбран гиперболический тангенс.

Из рисунка видно, что увеличение числа сенсорных нейронов способствует увеличению точности прогноза и если для сети с 64-мя нейронами горизонт предсказания может быть принят равным 10 (в крайнем случае – 20), рис. 1 *a, b*, то в случае 200-х нейронов, рис. 1 *c, d*, он увеличивается до 100. Конечно, многое здесь зависит не столько от числа сенсорных



нейронов, сколько от соотношения периода или почти периода сигнала и числа выборок сигнала на этом периоде, используемых при прогнозировании. Так, при увеличении частоты сигнала (при меньшей величине его периода) хорошие результаты можно получить и по 64 выборкам сигнала.

Увеличение числа нейронов скрытого слоя также ведёт к повышению точности прогноза, но только в заметно меньшей степени. Конечно, в этом случае многое зависит от выбранной процедуры обучения, точнее, от способа формирования обучающих образов.

Заключение

Существующие методы прогнозирования временных рядов и непрерывных процессов характеризуются большим разнообразием, что в свою очередь, обусловлено большим разнообразием задач. Так, велико число работ по методам прогноза, основанным на идеях декомпозиции процессов по эмпирическим модам (метод EMD). Однако данный подход приме-

ним, как правило, только при условии, что наблюдаемый процесс представлен небольшим числом гармонических составляющих, модулированных по амплитуде и фазе.

Метод параметрической регрессии, используемый в ряде задач аппроксимации [31, 32], обладает существенным недостатком, заключающимся в возможности возникновения неконтролируемой ошибки на интервале прогнозирования, особенно при высоком порядке прогнозирующей функции.

Особый интерес в плане прогнозирования связан с представлением временных рядов моделями авторегрессии, нелинейными системами Вольтерры и нейронными сетями. Несмотря на обилие работ в этой области, задача прогнозирования процессов на несколько шагов вперед представлена недостаточно. Обычно всё ограничивается прогнозом на один шаг.

В настоящей работе показано, что в рамках модели авторегрессии прогноз на несколько шагов вперед зависит от коэффициентов модели нелинейным образом. При этом задачу про-

гноза, состоящую в нахождении коэффициентов модели посредством минимизации целевой функции, можно решить только итерационным методом.

Выводы, полученные для модели авторегрессии, можно распространить, в принципе, и на модели в форме нелинейных систем Вольterra. Потребность в подобных системах обычно проистекает из нелинейности источников, порождающих наблюдаемый процесс. Полученные уравнения, основанные на таких системах, открывают новые возможности по решению задач прогнозирования.

К классу нелинейных систем относятся также и искусственные нейронные сети. Возможно, наибольший интерес в плане прогноза представляют рекуррентные и другие сети с обратными связями. Тем не менее, в настоящей работе всё внимание сосредоточено на нейронных сетях прямого распространения. В работе получены уравнения функционирования и обучения в матричной форме, получен алгоритм обратной подстановки, с помощью которого можно увеличить глубину прогноза. Результаты моделирования показали эффективность разработанных алгоритмов.

Литература

1. Кравцов Ю.Н. Случайность, детерминированность, предсказуемость // УФН, т. 158, вып. 1, 1989. С. 93 – 122.
2. Кравцов Ю.Н. Фундаментальные и практические пределы предсказуемости. В сб. ст. Пределы предсказуемости. Под ред. Ю.Н. Кравцова. – М.: ЦентрКом, 1997. – 247 с.
3. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. В 2-х выпусках. – М.: Мир, 1974. Вып. 1 – 408 с., Вып. 2 – 199 с.
4. Маевский В.В., Харин Ю.С. Робастность регрессионного прогнозирования при наличии функциональных искажений модели // АиТ, №11, 2012. – С. 118 – 137.
5. Huang N.E., Shen Z., Long S.R., Wu M.C., Shin H.H., Zheng Q., Yen N.-C., Tung C.C., Liu H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis // Proc. R. Soc. Lond. A (1998), 454, p. 903 – 995.
6. Huang N.E., Wu Z. A review on Hilbert-Huang transform: method and its applications to geophysical studies // Reviews of Geophysics, 46, RG2006 / 2008, p. 1 – 23.
7. Chu P.C., Fan C., Huang N. Derivative-optimized empirical mode decomposition for the Hilbert-Huang transform // Journal of Computational and Applied Mathematics, 259, 2014, p. 57–64.
8. Курбацкий В.Г., Сидоров Д.Н., Спиряев В.А., Томин Н.В. О нейросетевом подходе к прогнозированию нестационарных временных рядов на основе преобразования Гильберта-Хуанга // АиТ, №7, 2011. – С. 58 – 68.
9. Вайнштейн Л.А., Вакман Д.Е. Разделение частот в теории колебаний и волн. – М.: Наука, 1983. – 288 с.
10. Flandrin P., Rilling G., Goncalves P. Empirical mode decomposition as a filter bank // IEEE Signal Proc. Lett., 2004, 11, p. 112 – 114.
11. Angelosante D., Giannakis G.B., Sidiropoulos N.D. Sparse parametric models for robust nonstationary signal analysis // IEEE Signal Processing Magazine, November, 2013, p. 64 – 73.
12. Гроп Д. Методы идентификации систем. – М.: Мир, 1979. – 304 с.
13. Foster P., Klapuri A., Dixon S. A method for identifying repetition structure in musical audio based on time series prediction // 20th European Signal Processing Conference (EUSIPCO 2012), 2012, 1299 – 1303.
14. Ni H., Yin H. A self-organising mixture autoregressive network for FX time series modelling and prediction // Neurocomputing, 72, 2009, p. 3529 – 3537.
15. Richard C., Bermudez J.C.M., Honeine P. Online prediction of time series data with kernels // IEEE Transactions on Signal Processing, Vol. 57, No.3, 2009, p. 1058 – 1067.
16. Торгашов А.Ю. Робастно – стабилизирующие горизонты регулятора на основе прогнозирующей модели с усеченной импульсной переходной функцией // АиТ, № 7, 2007. – С. 90 – 102.
17. Li Y., Kashiwagi H. High-order Volterra model predictive control and its application to a nonlinear polymerization process // International Journal of Automation and Computing 2, 2005, p. 208 – 214.
18. Samakee C., Phukpattaranont P. Application of MISO Volterra series for modeling subharmonic of ultrasound contrast agent // International Journal of Computer and Electrical Engineering, Vol. 4, No. 4, 2012, p. 445 – 451.
19. Mandic D.P., Chambers J.A. Recurrent neural networks for prediction. Learning algorithms. Architecture and stability. – Chichester: John Wiley & Sons, 2001. – 285 p.
20. Bouchachia A. Radial basis function nets for time series prediction // International Journal of Computational Intelligence Systems, Vol. 2, No. 2, 2009, p. 147 – 157.
21. Yee P., Haykin S. A dynamic regularized radial basis function network for nonlinear, nonstationary

time series prediction // IEEE Transaction on Sign. Processing, Vol. 47, No. 9, 1999, p. 2503–2521.

22. Нестационарные системы автоматического управления: анализ, синтез и оптимизация / Под ред. К.А. Пупкова и Н.Д. Егупова. – М.: Издательство МГТУ им. Н.Э. Баумана, 2007. – 632 с.

23. Куркин О.М. Исследование алгоритмов гарантирующего оценивания в задачах прогнозирования и интерполяции случайных процессов // АиТ, № 4, 2001. - С. 67 – 79.

24. Буцев А.В., Первозванский А.А. Локальная аппроксимация на искусственных нейросетях // АиТ, № 9, 1995. - С. 127 – 136.

25. Виллемс Я. От временного ряда к линейной системе. В сборнике статей Теория систем. Математические методы и моделирование. (Серия: Математика. Новое в зарубежной науке. Вып. 44. Под ред. А.Н. Колмогорова и С.П. Новикова.) – М.: Мир, 1989. – 384 с.

26. Современные методы идентификации систем. / Под ред. П. Эйкхоффа. – М.: Мир, 1983. – 400 с.

27. Льюнг Л. Идентификация систем. Теория для пользователя. – М.: Наука, 1991. – 432 с.

28. Гельфонд А.О. Исчисление конечных разностей. – М.: Наука, 1967. – 376 с.

29. Пупков К.А., Капалин В.И., Ющенко А.С. Функциональные ряды в теории нелинейных систем. – М.: Наука, 1976. – 448.

30. Ермолаев В.А. Вариационный синтез входных сигналов нелинейных фильтров по их отклику в фиксированный момент времени // РЭ, 1987, № 6, с. 1199 – 1206.

31. Ермолаев В.А., Карасев О.Е., Кропотов Ю.А. Метод интерполяционной фильтрации в задачах обработки речевых сигналов во временной области // Вестник компьютерных и информационных технологий, № 7, 2008. - С. 12 – 17.

32. Кропотов Ю.А., Ермолаев В.А. Об адаптивном алгоритме наименьших квадратов в задачах компенсации акустического эха // Радиотехнические и телекоммуникационные системы, №2(6), 2012. - С. 40-44.

Поступила 30 ноября 2015 г.

English

Predictive methods of time series and continuous processes

Valery Andreevich Yermolaev – Candidate of Technical Sciences Associate Professor Murom Institute (branch) Vladimir State University named after Alexander and Nickolay Stoletov.

E-mail: valeermolaev@yandex.ru.

Address: Orlovskaya st., 23, Murom, 602264.

Abstract: Possibilities of predictive algorithms of various classes are defined. Limited possibilities of processes decomposition methods into empiric modes and parametrical methods of prediction based on presentation of time series via the generalized polynom according to the principle of linearly independent functions are indicated. Predictive algorithms based on autoregression models are obtained. It is illustrated that predictive results for two and greater number of steps forward are nonlinear functions of autoregression model coefficients. Recurrent equations defining explicit dependence of prediction evaluations on model coefficients are obtained. Task is set to define prediction evaluations via minimization of loss function - square of evaluations tolerance rate of existing time series. Method of a predictive algorithm generalization via autoregression linear model onto nonlinear models - models rendered by functional series or Volterra systems is obtained. The predictive algorithm of time series for several steps forward via artificial feedforward neural networks is obtained. Two predictive algorithms are proposed. The first one is based on minimization of loss function rendered via square of predictive vector difference rate at network output and expect data vector. The second approach presupposes inverse substitution procedure when shift of vector components takes place and that is exit of sensor layer with subsequent occupation of vacant place by the network outputs while preserving its weight factor values. Inverse substitution procedure is rendered via difference equation. Modeling of operation algorithms and neural networks training showed their efficiency. The evaluation of dependence of predictive horizon on sensor layer sizes and the compliance of its sizes almost to the period of the observed time series is established.

Key words: prediction, predictive horizon, autoregression, functional series, Volterra nonlinear system, neural network.

References

1. Kravtsov Yu.N. Randomness, determinancy, predictability//UFN, v. 158, iss. 1, 1989, P. 93 - 122.
2. Kravtsov Yu.N. Fundamental and practical limits of predictability. In art. collect. Predictability limits. Ed. by Yu.N. Kravtsov. - М.: Tsentrkom, 1997. - 247 p.

3. George Box, Gwilym Jenkins Time Series Analysis: Forecasting and Control. In 2 ed. - M.: Mir, 1974. Iss. 1-408 p., Iss. 2 - 199 p.
4. Mayevsky V. V., Kharin Yu.S. Regressive forecasting robustness if any model functional distortions. - AIT, No. 11, 2012, P.118 - 137.
5. Huang N.E., Shen Z., Long S.R., Wu M.C., Shin H.H., Zheng Q., Yen N. - C., Tung C.C., Liu H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. - Proc. R. Soc. Lond. A (1998), 454, p. 903 - 995.
6. Huang N.E., Wu Z. A review on Hilbert-Huang transform: method and its applications to geophysical studies. - Reviews of Geophysics, 46, RG2006/2008, p. 1 - 23.
7. Chu P.C., Fan C., Huang N. Derivative-optimized empirical mode decomposition for the Hilbert-Huang transform. - Journal of Computational and Applied Mathematics, 259, 2014, p. 57-64.
8. Kurbatsky V. G., Sidorov D. N., Spiryaev V.A., Tomin N. V. On neural network approach to predicting non-stationary time series based on Hilbert-Huang transform. - AIT, No. 7, 2011, P. 58 - 68.
9. L. A. Weinstein, D. E. Vakman, Frequencies separation in theory of oscillations and waves. - M.: Nauka, 1983. - 288 p.
10. Flandrin P., Rilling G., Goncalves P. Empirical mode decomposition as a filter bank//IEEE Signal Proc. Lett., 2004, 11, p. 112 - 114.
11. Angelosante D., Giannakis G.B., Sidiropoulos N.D. Sparse parametric models for robust nonstationary signal analysis. - IEEE Signal Processing Magazine, November, 2013, p. 64 - 73.
12. Grop D. Methods of systems identification of. - M.: Mir, 1979. - 304 p.
13. Foster P., Klapuri A., Dixon S. A method for identifying repetition structure in musical audio based on time series prediction. - 20th European Signal Processing Conference (EUSIPCO 2012), 2012, 1299 - 1303.
14. Ni H., Yin H. A self-organizing mixture autoregressive network for FX time series modeling and prediction. - Neurocomputing, 72, 2009, p. 3529 - 3537.
15. Richard C., Bermudez J.C.M., Honeine P. Online prediction of time series data with kernels. - IEEE Transactions on Signal Processing, Vol. 57, No. 3, 2009, p. 1058 - 1067.
16. Torgashov A.Yu. Robustly stabilizing regulator level axis based on predictive model with trimmed impulsive admittance function. - AIT, No. 7, 2007, p. 90 - 102.
17. Li Y., Kashiwagi H. High-order Volterra model predictive control and its application to a nonlinear polymerization process. - International Journal of Automation and Computing 2, 2005, p. 208 - 214.
18. Samakee C., Phukpattaranont P. Application of MISO Volterra series for modeling subharmonic of ultrasound contrast agent. - International Journal of Computer and Electrical Engineering, Vol. 4, No. 4, 2012, p. 445 - 451.
19. Mandic D.P., Chambers J.A. Recurrent neural networks for prediction. Learning algorithms. Architecture and stability. - Chichester: John Wiley & Sons, 2001. - 285 p.
20. Bouchachia A. Radial basis function nets for time series prediction//International Journal of Computational Intelligence Systems, Vol. 2, No. 2, 2009, p. 147 - 157.
21. Yee P., Haykin S. A dynamic regularized radial basis function network for nonlinear, nonstationary time series prediction. - IEEE Transaction on Sign. Processing, Vol. 47, No. 9, 1999, p. 2503-2521.
22. Non-stationary automatic control systems: the analysis, synthesis and optimization. - Ed. by K.A. Pupkov and N. D. Egupov. - M.: Bauman MSTU 2007. - 632 p.
23. Kurkin O. M. Research of ensured estimation algorithms for tasks of prediction and interpolation of accidental processes. - AIT, No. 4, 2001, P. 67 - 79.
24. Butsev A.V., Pervozvansky A.A. Local approximation on artificial neuronets. - AIT, No. 9, 1995, P. 127 - 136.
25. Willems J. From time series to linear system. Articles collection Mathematical methods and modeling. (Ser.: Mathematics. New in foreign science. iss. 44. Ed. by A.N. Kolmogorov and S. P. Novikov.) - M.: Nauka, 1989. - 384 p.
26. Trends and progress in system identification. / Ed. by P. Eykhoff. - M.: Mir, 1983. - 400 P.
27. Lennart Ljung. System Identification. Theory for user. - M.: Nauka, 1991. - 432 P.
28. A.O. Gelfond. Calculus of finite differences. - M.: Nauka, 1967. - 376 P.
29. Pupkov K.A., Kapalin V. I., Yushchenko A.S. Functional Series in theory of nonlinear systems. - M.: Nauka, 1976. - 448.
30. Yermolaev V.A. Variation synthesis of input signals of nonlinear filters according to their response at a fixed time. - RE, 1987, No. 6, P. 1199 - 1206.
31. Yermolaev V.A., Karasyov O. E., Kropotov Yu.A. Method of interpolation filtering for processing voice signals in the time domain. - Vestnik kompyuternykh i informatsionnykh tekhnology, No. 7, 2008, P. 12 - 17.
32. Kropotov Yu.A., Yermolaev V.A. On adaptive algorithm of the least squares for tasks of compensation of acoustic echo. - Radiotekhnicheskiye i telekommunikatsionnyye sistemy, No. 2(6), 2012, P.40-44.