

АДАПТИВНОЕ ОБУЧЕНИЕ ВЭЙВЛЕТ-МАШИНЫ ОПОРНЫХ ВЕКТОРОВ НАИМЕНЬШИХ КВАДРАТОВ

Е. А. Винокурова, И. П. Плисс, К. А. Шкуро, А. А. Харченко

Харьковский национальный университет радиоэлектроники

просп. Ленина, 14, г. Харьков, 61166, Украина. E-mail: vinokurova@kture.kharkov.ua, pliss@kture.kharkov.ua

Своеобразным гибридом нейронных сетей, обучение которых основывается как на оптимизации, так и на памяти, являются машины опорных векторов (SVM), чья архитектура совпадает с радиально-базисной нейронной сетью и обобщенной регрессионной нейронной сетью, синаптические веса определяются в результате решения задачи нелинейного программирования, а центры активационных функций устанавливаются по принципу «нейроны в точках данных». И хотя SVM-сети обладают целым рядом несомненным преимуществ, их обучение с вычислительной точки зрения представляется достаточно трудоемким, поскольку связано с решением задач нелинейного программирования высокой размерности. В связи с этим в качестве альтернативы SVM были предложены машины опорных векторов наименьших квадратов (LS-SVM), обучение которых сводится к решению систем линейных уравнений, что гораздо проще вычислительной точки зрения и может быть реализовано в on-line режиме. Вэйвлет-аналогом традиционной SVM является вэйвлет-машина опорных векторов (WSVM), в которой многомерные ядерные функции активации заменяются одномерными адаптивными вэйвлет-функциями. И хотя WSVM обладает большими возможностями по сравнению с традиционными SVM, их обучение с вычислительной точки зрения связано с реализацией достаточно сложных процедур, что, естественно, ограничивает их возможности для решения задач реального времени. В связи с этим представляется целесообразной разработка достаточно простых вэйвлет-нейро-фаззи-систем, реализующих идеи обучения, основанного на минимизации эмпирического риска и ориентированных на обработку информации в on-line-режиме. В качестве базового элемента таких систем нами принят вэйвлет-нейрон, характеризующийся высокими аппроксимирующими свойствами, простотой, высокой скоростью обучения и возможностью выявлять скрытые зависимости в обрабатываемых данных.

Ключевые слова: машины опорных векторов наименьших квадратов, вэйвлет-нейрон, прогнозирование, нелинейные нестационарные временные ряды, адаптивная вэйвлет-функция

АДАПТИВНЕ НАВЧАННЯ ВЕЙВЛЕТ-МАШИНИ ОПОРНИХ ВЕКТОРІВ НАЙМЕНЬШИХ КВАДРАТІВ

О. А. Винокурова, І. П. Плісс, К. О. Шкуро, О. О. Харченко

Харківський національний університет радіоелектроніки

просп. Леніна, 14, м. Харків, 61166, Україна. E-mail: vinokurova@kture.kharkov.ua, pliss@kture.kharkov.ua

Своєрідним гібридом нейронних мереж, навчання яких базується як на оптимізації, так і на пам'яті, є машини опорних векторів (SVM), чия архітектура збігається з радіально-базисною нейронною мережею і узагальненою регресійною нейронною мережею, синаптичні ваги визначаються в результаті вирішення задачі нелінійного програмування, а центри активационних функцій встановлюються по принципу «нейрони в точках даних». І хоча SVM-мережі мають цілу низку безсумнівних переваг, їх навчання з обчислювальної точки зору є достатньо трудомістким, оскільки пов'язано з вирішенням задач нелінійного програмування високої розмірності. У зв'язку з цим як альтернативу SVM були запропоновані машини опорних векторів найменших квадратів (LS-SVM), навчання яких зводиться до вирішення систем лінійних рівнянь, що набагато простіше з обчислювальної точки зору і може бути реалізовано в on-line режимі. Вейвлет-аналогом традиційної SVM є вейвлет машина опорних векторів (WSVM), у якій багатовимірні ядерні функції активації замінені одновимірними адаптивними вейвлет-функціями. І хоча WSVM має великі можливості в порівнянні з традиційними SVM, їх навчання з обчислювальної точки зору пов'язано з реалізацією достатньо складних процедур, що, природно, обмежує їх можливості для вирішення задач реального часу. У зв'язку з цим є доцільною розробка достатньо простих вейвлет-нейро-фаззи-систем, що реалізують ідеї навчання, заснованого на мінімізації критерію емпіричного ризику і орієнтованих на обробку інформації в on-line-режимі. Як базовий елемент таких систем нами прийнято вейвлет-нейрон, що характеризується високими апроксимуючими властивостями, простотою, високою швидкістю навчання і можливістю виявляти приховані залежності в даних, що оброблюються.

Ключові слова: машини опорних векторів, вейвлет-нейрон, прогнозування, нелінійні нестационарні часові ряди, адаптивна вейвлет-функція.

АКТУАЛЬНОСТЬ РАБОТЫ. На сегодняшний день системы вычислительного интеллекта получили широкое распространение для решения обширного класса задач, связанных с обработкой информации, представленной в самых различных формах – от простейших таблиц «объект-свойство» до сложных нестационарных многомерных стохастических временных рядов. Системы вычислительного интеллекта, такие как искусственные нейронные сети, нейро-фаззи системы, вэйвлет-системы обладают высоким быстродействием, универсальными аппроксимирующими свойствами, выявлением локальных особенностей [1–7].

Существующие подходы для своего обучения требуют больших объемов обучающей выборки, причем этот объем исходной выборки должен быть как минимум на два порядка больше, чем количество оцениваемых параметров таких систем. Но, к сожалению, такая ситуация возникает далеко не всегда, особенно при решении практических задач. Во многих практических задачах, например, в задачах медицинской диагностики, прогнозирования финансовых показателей и т.д., объемы обучающей выборки совершенно недостаточно для того, чтобы построить и обучить эффективную систему вычислительного интеллекта.

Решение задач в таких ситуаціях з допомогою різних методів теорії ідентифікації [8] було не ефективним, в зв'язі з чим Вапником В.Н. [9–12] был предложен, так называемый, метод минимизации эмпирического риска и на его основе построена машина опорных векторов.

Стандартная машина опорных векторов – это вычислительная система, которая минимизирует эмпирический риск, но с точки зрения практической реализации это достаточно сложная система, поскольку связана с решением на каждом такте задачи нелинейного программирования, причем высокой размерности с ограничениями типа неравенств. Поэтому в такой ситуации, идея была замечательной, однако ее реализация в online режиме была не продуктивна [2].

Поэтому была предложена модификация этой системы, получившая название LS-SVM (машина опорных векторов наименьших квадратов). Здесь были изменены условия так, что на каждом шаге надо было решать задачу квадратичного программирования и наличие ограничений типа равенств, а уже это создавало предпосылки к нейросетевой реализации этого подхода. Но все равно эта машина оказалась достаточно сложна и поэтому в дальнейшем были предприняты попытки улучшения, во-первых, быстродействия, упрощения вычислительной реализации и, во-вторых, сокращения объема обучающей выборки [13].

На основе нечеткой SVM [14, 15] была создана вэйвлет-машина опорных векторов наименьших квадратов и по сравнению со всеми предыдущими вариантами она обладала, конечно же, улучшенными аппроксимирующими свойствами, но при этом резко усложнилась численная реализация [16, 17]. В основе машины LS-SVM лежит идеология РБНС [18], где радиально- базисные функции заменены многомерными вэйвлетами. Для того чтобы упростить программную реализацию и сократить количество настраиваемых параметров и повысить быстродействие в работе предлагается использовать как базовую архитектуру вэйвлет-нейрон [19–21] (который показал высокие аппроксимирующие свойства, высокое быстродействие). Но имеющиеся алгоритмы обучения не позволяют обучить вэйвлет-нейрон в условиях малой и сверхмалой выборки.

Альтернативой обучению, основанному на оптимизации, является обучение, основанное на памяти и связанное с концепцией «нейроны в точках данных». Наиболее характерным представителем нейронных сетей, обучение которых основано на этом принципе, являются обобщенные регрессионные нейронные сети (GRNN), однако они решают задачу интерполяции, а не аппроксимации, что существенно усложняет их использование в задачах обработки «зашумленных» данных.

В связи с этим представляется целесообразной разработка достаточно простых вэйвлет-нейрофаззи систем, реализующих идеи обучения, основанного на минимизации эмпирического риска, ориентированных на обработку информации в online-режиме. В качестве базового элемента таких систем может быть принят вэйвлет-нейрон, характеризую-

щийся высокими аппроксимирующими и экстраполирующими свойствами.

Архитектура вэйвлет-нейрона. Рассмотрим архитектуру вэйвлет-нейрона, приведенную на рис. 1 [21].

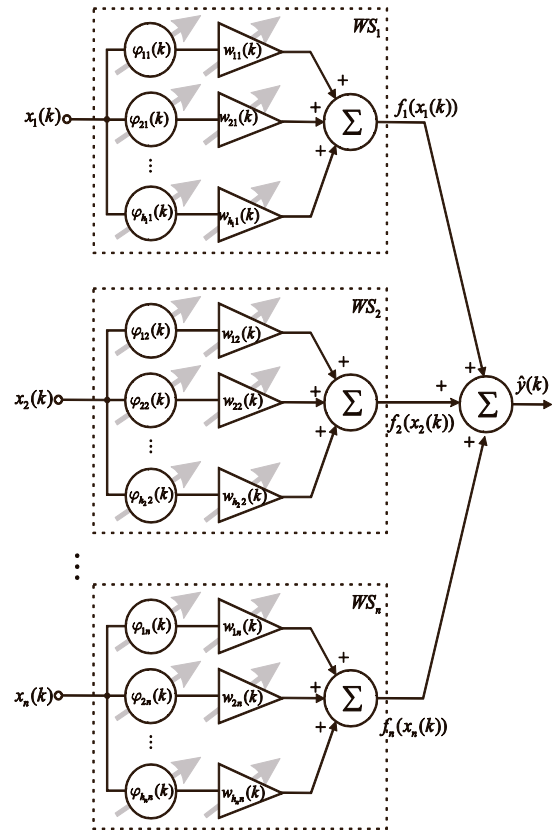


Рисунок 1 – Архитектура вэйвлет-нейрона

Как видно, вэйвлет-нейрон достаточно близок по конструкции к стандартному n – входовому формальному нейрону, однако вместо обычных настраиваемых синаптических весов содержит вэйвлет-синапсы $WS_i, i = 1, 2, \dots, n$ обучаемыми параметрами которых являются не только веса $w_{ji}(k)$, но и параметры центра, ширина и форма адаптивной вэйвлет-функции активации $\varphi_{ji}(x_i(k))$.

При подаче на вход вэйвлет-нейрона векторного сигнала

$$x(k) = (x_1(k), x_2(k), \dots, x_n(k))^T$$

(здесь $k = 1, 2, \dots$ - текущее дискретное время) на его выходе появляется скалярное значение.

$$\hat{y}(k) = \sum_{i=1}^n f_i(x_i(k)) = \sum_{i=1}^n \sum_{j=1}^{h_i} w_{ji}(k) \varphi_{ji}(x_i(k)) \cdot \quad (1)$$

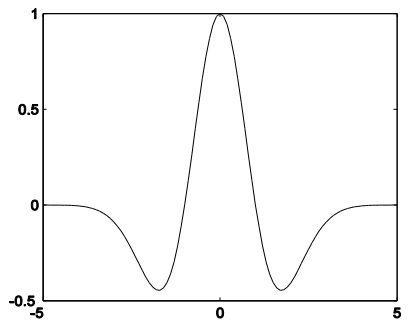
В качестве функций активации в составе вэйвлет-нейрона могут использоваться любые семейства аналитических четных вэйвлетов. Введем в рассмотрение математическую модель одномерной адаптивной вэйвлет-функции активации-принадлежности, которую синтезируем на основе вэйвлета "Мексиканская шляпа" и имеющую вид [22, 23]

$$\varphi_{ji}(x_i(k)) = (1 - \alpha_{ji}(k)t_{ji}^2(k)) \exp(-t_{ji}^2(k)/2), \quad (2)$$

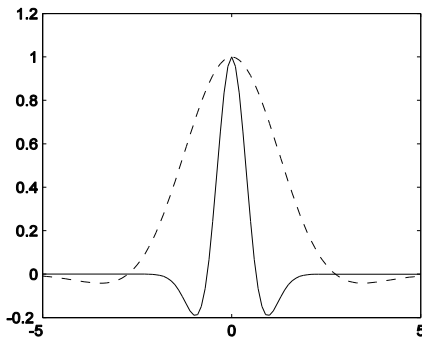
где $t_{ji}(k) = (x_i(k) - c_{ji}(k))/\sigma_{ji}(k)$ $c_{ji}(k)$ параметр, определяющий положение центра функции; $\sigma_{ji}(k)$ параметр, определяющий ширину функции; $\alpha_{ji}(k)$ параметр, определяющий форму вэйвлет-функции.

Уточняемый параметр α_{ji} позволяет настраивать форму адаптивной вэйвлет-функции активации-принадлежности в процессе обучения сети, при этом если $\alpha_{ji} = 0$ получаем Гауссову функцию активации, при $\alpha_{ji} = 1$ получаем вэйвлет-функцию "Mexican Hat", а при $0 < \alpha_{ji} < 1$ гибридную функцию активации-принадлежности.

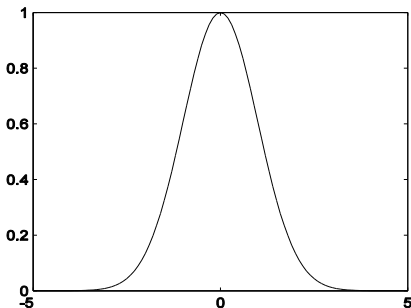
На рис. 2 представлены адаптивные вэйвлет-функции активации при разных параметрах α и σ .



а)



б)



в)

Рисунок 2 – Адаптивная вэйвлет-функция:

а – $\alpha = 1, \sigma = 1$; б – пунктирная линия,

$\alpha = 0.3, \sigma = 1.5$; сплошная линия $\alpha = 0.6, \sigma = 0.5$;

в – $\alpha = 0, \sigma = 1$

Задача обучения состоит в том, чтобы рассчитать на каждой итерации k синаптические веса $w_{ji}(k)$,

центры $c_{ji}(k)$, ширины $\sigma_{ji}^{-1}(k)$ и параметры формы $\alpha_{ji}(k)$ вэйвлет-функций принадлежности-активации, которые будут оптимизировать критерий качества.

Алгоритм обучения всех параметров вэйвлет-нейрона. В качестве критерия обучения вэйвлет-нейрона используется традиционная квадратичная функция ошибки, которую в обозначениях вэйвлет-нейрона можно записать следующим образом:

$$E(k) = \frac{1}{2} (y(k) - \hat{y}(k))^2 = \frac{1}{2} \left(y(k) - \sum_{i=1}^n \sum_{j=1}^{h_i} w_{ji}(k) \varphi_{ji}(x_i(k)) \right)^2, \quad (3)$$

где $y(k)$ – внешний обучающий сигнал.

Введя ряд обозначений

$$\varphi_i(x_i(k)) = (\varphi_{1i}(x_i(k)), \dots, \varphi_{h_i i}(x_i(k)))^T,$$

$$w_i(k) = (w_{1i}(k), \dots, w_{h_i i}(k))^T,$$

$$c_i(k) = (c_{1i}(k), \dots, c_{h_i i}(k))^T,$$

$$\sigma_i^{-1}(k) = (\sigma_{1i}^{-1}(k), \dots, \sigma_{h_i i}^{-1}(k))^T,$$

$$\alpha_i(k) = (\alpha_{1i}(k), \dots, \alpha_{h_i i}(k))^T, \quad t_i(k) = (t_{1i}(k), \dots, t_{h_i i}(k))^T$$

можно записать алгоритм обучения в виде:

$$\begin{cases} w_i(k+1) = w_i(k) + (e(k)J_i^w(k))/r_i^w(k), \\ c_i(k+1) = c_i(k) + (e(k)J_i^c(k))/r_i^c(k), \\ \sigma_i^{-1}(k+1) = \sigma_i^{-1}(k) + (e(k)J_i^\sigma(k))/r_i^\sigma(k), \\ \alpha_i(k+1) = \alpha_i(k) + (e(k)J_i^\alpha(k))/r_i^\alpha(k), \end{cases} \quad (4)$$

$$\begin{cases} r_i^w(k+1) = \beta r_i^w(k) + \|J_i^w(k)\|^2, \\ r_i^c(k+1) = \beta r_i^c(k) + \|J_i^c(k)\|^2, \\ r_i^\sigma(k+1) = \beta r_i^\sigma(k) + \|J_i^\sigma(k)\|^2, \\ r_i^\alpha(k+1) = \beta r_i^\alpha(k) + \|J_i^\alpha(k)\|^2, \end{cases} \quad (5)$$

где $0 \leq \beta \leq 1$ – параметр забывания,

$$J_i^w(k) = (J_{1i}^w(k), \dots, J_{h_i i}^w(k))^T,$$

$$J_i^c(k) = (J_{1i}^c(k), \dots, J_{h_i i}^c(k))^T,$$

$$J_i^\sigma(k) = (J_{1i}^\sigma(k), \dots, J_{h_i i}^\sigma(k))^T,$$

$$J_i^\alpha(k) = (J_{1i}^\alpha(k), \dots, J_{h_i i}^\alpha(k))^T,$$

$$J_{ji}^w(k) = (1 - \alpha_{ji}(k)t_{ji}^2(k)) \exp(-t_{ji}^2(k)/2),$$

$$J_{ji}^c(k) = w_{ji}(k) \sigma_{ji}^{-1}(k) \times$$

$$\times ((2\alpha_{ji}(k) + 1)t_{ji}(k) - \alpha_{ji}(k)t_{ji}^3(k)) \exp(-t_{ji}^2(k)/2),$$

$$J_{ji}^{\sigma}(k) = w_{ji}(k)(x_i(k) - c_{ji}(k)) \times \\ \times \left((2\alpha_{ji}(k) + 1)t_{ji}(k) - \alpha_{ji}(k)t_{ji}^3(k) \right) \exp(-t_{ji}^2(k)/2), \\ J_{ji}^{\alpha}(k) = e(k)w_{ji}(k)t_{ji}^2(k) \exp(-t_{ji}^2(k)/2).$$

Несложно видеть, что при $\beta = 1$ метод (4), (5) приобретает свойства стохастической аппроксимации адаптивного алгоритма идентификации Гудвина-Рэмеджа-Кэйнеса [24], а при $\beta = 0$ имеет форму популярного в теории искусственных нейронных сетей алгоритма Уидроу-Хоффа.

Как видно, использование модифицированных квази-ньютоновских методов обучения практически не усложняет численную реализацию процедур настройки вэйвлет-синапсов, обеспечивая при этом повышение их скорости сходимости.

Обучение вэйвлет машины опорных векторов наименьших квадратов. В случае, когда обрабатываемый сигнал описывается короткой или сверхкороткой выборкой, метод обучения, предложенный выше, не успевает обучить все параметры сети. Поэтому на первый план выходят методы обучения синаптических весов на основе минимизации эмпирического риска.

Вводя в рассмотрение $(nh \times 1)$ – вектор адаптивных вэйвлет функций активации $\varphi(x(k)) = (\varphi_{11}(x_1(k)), \dots, \varphi_{h1}(x_1(k)), \varphi_{12}(x_2(k)), \dots, \varphi_{h2}(x_2(k)), \dots, \varphi_{in}(x_n(k)), \dots, \varphi_{hn}(x_n(k)))^T$,

и соответствующий ему вектор синаптических весов вэйвлет-нейрона

$$w = (w_{11}, \dots, w_{h1}, w_{12}, \dots, w_{h2}, \dots, w_{in}, \dots, w_{hn})^T. \quad (7)$$

можно переписать преобразование (1) реализуемое вэйвлет-нейроном в форме

$$\hat{y}(k) = w^T \varphi(x(k)). \quad (8)$$

Обучение вэйвлет-нейрона с использованием опорных векторов наименьших квадратов (LS-SVM-WN) сводится к оптимизации квадратичного критерия

$$E(N) = \frac{1}{2} \|w\|^2 + \frac{\gamma}{2} \sum_{k=1}^N e^2(k) \quad (9)$$

с учетом ограничений в виде системы из N линейных уравнений

$$y(k) = w^T \varphi(x(k)) + e(k), \quad (10)$$

где $\gamma > 0$ – параметр регуляризации.

Оптимизация критерия (9) без учета ограничений (10) ведет к выражению

$$w(N) = \left(\sum_{k=1}^N \varphi(x(k)) \varphi^T(x(k)) + \gamma^{-1} I \right)^{-1} \times \\ \times \sum_{k=1}^N \varphi(x(k)) y(k), \quad (11)$$

весьма близкому по сути к ридж-оценки (смещенной оценки наименьших квадратов), где $I - (nh \times nh)$ – единичная матрица.

Для учета системы ограничений (10) введем в рассмотрение функцию Лагранжа

$$L(w, e(k), \lambda(k)) = E(k) + \\ + \sum_{k=1}^N \lambda(k) (y(k) - w^T \varphi(x(k)) - e(k)) = \\ = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{k=1}^N e^2(k) + \\ + \sum_{k=1}^N \lambda(k) (y(k) - w^T \varphi(x(k)) - e(k)) \quad (12)$$

(здесь $\lambda(k)$ – N неопределенных множителей Лагранжа) и систему уравнений Каруша–Куна–Таккера

$$\begin{cases} \nabla_w L(w, e(k), \lambda(k)) = w - \sum_{k=1}^N \lambda(k) \varphi(x(k)) = \vec{0}_N, \\ \frac{\partial L(w, e(k), \lambda(k))}{\partial e(k)} = \gamma e(k) - \lambda(k) = 0, \\ \frac{\partial L(w, e(k), \lambda(k))}{\partial e(k)} = y(k) - w^T \varphi(x(k)) - e(k) = 0, \end{cases} \quad (13)$$

где $\vec{0}_N - (N \times 1)$ – вектор, образованный нулями.

Решение системы (13) имеет вид:

$$\begin{cases} w(N) = \sum_{k=1}^N \lambda(k) \varphi(x(k)), \\ \lambda(k) = \gamma e(k), \\ y(k) = w^T(N) \varphi(x(k)) + e(k) \end{cases} \quad (14)$$

или в матричной форме

$$(\gamma^{-1} I_{NN} + \Omega_{NN}) \begin{pmatrix} \lambda(1) \\ \vdots \\ \lambda(N) \end{pmatrix} = \begin{pmatrix} y(1) \\ \vdots \\ y(N) \end{pmatrix}. \quad (15)$$

(здесь $I_{NN} - (N \times N)$ – единичная матрица) или

$$(\gamma^{-1} I_{NN} + \Omega_{NN}) \Lambda_N = Y_N \quad (16)$$

(здесь $\Omega_{NN} = \{\Omega_{pq} = \varphi^T(x(p)) \varphi(x(q))\}$ $p = 1, 2, \dots, N$; $q = 1, 2, \dots, N$ откуда

$$\Lambda_N = (\gamma^{-1} I_{NN} + \Omega_{NN})^{-1} Y_N. \quad (17)$$

Тогда выходной сигнал вэйвлет-нейрона имеет вид

$$\hat{y}(x) = \left(\sum_{k=1}^N \lambda(k) \varphi(x(k)) \right)^T \varphi(x). \quad (18)$$

Легко заметить, что с точки зрения нейроматематики (теории обучения искусственных нейронных сетей) и теории машин опорных векторов (минимизация эмпирического риска), предложенная LS-SVM на основе вэйвлет-нейрона гораздо проще в реализации, характеризуется высоким быстродействием, требует малых объемов обучающей выборки.

Если же данные на обработку поступают последовательно, процесс обучения LS-SVM-NFN следует организовать в online-режиме. При этом при поступлении пары $x(N+1), y(N+1)$ выражение (18) приобретает вид

$$\hat{y}(x) = \left(\sum_{k=1}^N \lambda(k) \varphi(x(k)) + \lambda(N+1) \varphi(x(N+1)) \right)^T \varphi(x) \quad (19)$$

или в матричной форме

$$(\gamma^{-1} \mathbf{I}_{N+1, N+1} + \Omega_{N+1, N+1}) \begin{pmatrix} \lambda(1) \\ \vdots \\ \lambda(N) \\ \text{-----} \\ \lambda(N+1) \end{pmatrix} = \begin{pmatrix} Y(1) \\ \vdots \\ Y(N) \\ \text{-----} \\ Y(N+1) \end{pmatrix}, \quad (20)$$

или

$$\begin{pmatrix} \Omega_{NN} & | & \omega_{N+1} \\ \text{-----} & & \text{-----} \\ \omega_{N+1}^T & | & \gamma^{-1} \end{pmatrix} \begin{pmatrix} \Lambda_N \\ \text{-----} \\ \lambda(N+1) \end{pmatrix} = \begin{pmatrix} Y_N \\ \text{-----} \\ y(N+1) \end{pmatrix}, \quad (21)$$

где

$$\omega_{N+1} = (\mu^T(x(1))\mu(x(N+1)), \mu^T(x(2))\mu(x(N+1)), \dots, \mu^T(x(N))\mu^T(x(N+1)))^T.$$

Из (22) следует

$$\Lambda_{N+1} = \begin{pmatrix} \Lambda_N \\ \text{-----} \\ \lambda(N+1) \end{pmatrix} = \begin{pmatrix} \Omega_{NN} & | & \omega_{N+1} \\ \text{-----} & & \text{-----} \\ \omega_{N+1}^T & | & \gamma^{-1} \end{pmatrix}^{-1} \begin{pmatrix} Y_N \\ \text{-----} \\ y(N+1) \end{pmatrix}. \quad (22)$$

Воспользовавшись далее формулой Фробениуса в виде [25]

$$M = \begin{pmatrix} A & | & B \\ - & - & - \\ C & | & D \end{pmatrix}, |D| \neq 0,$$

$$M^{-1} = \begin{pmatrix} A & | & B \\ - & - & - \\ C & | & D \end{pmatrix}^{-1} = \begin{pmatrix} K^{-1} & | & -K^{-1}BD^{-1} \\ \text{-----} & - & \text{-----} \\ -D^{-1}CK^{-1} & | & D^{-1} + D^{-1}CK^{-1}BD^{-1} \end{pmatrix},$$

$$K = A - BD^{-1}C,$$

можно записать

$$K = \Omega_{NN} - \omega_{N+1} \omega_{N+1}^T, \quad K^{-1} = (\Omega_{NN} - \gamma \omega_{N+1} \omega_{N+1}^T)^{-1}.$$

Несложно рассчитать $(N+1)$ -й множитель Лагранжа с помощью выражений

$$\begin{aligned} \lambda(N+1) &= -\gamma \omega_{N+1}^T K^{-1} Y_N + \\ &+ \gamma (1 + \gamma \omega_{N+1}^T K^{-1} \omega_{N+1}) y(N+1). \end{aligned} \quad (23)$$

Далее, воспользовавшись формулой Шермана-Мориссона обращения матриц, окончательно получаем

$$\begin{cases} K^{-1} = \Omega_{NN}^{-1} + \frac{\Omega_{NN}^{-1} \omega_{N+1} \omega_{N+1}^T \Omega_{NN}^{-1}}{1 - \omega_{N+1}^T \Omega_{NN}^{-1} \omega_{N+1}}, \\ \lambda(N+1) = 1 + \gamma \omega_{N+1}^T K^{-1} (\omega_{N+1} - Y_N). \end{cases} \quad (24)$$

Результаты моделирования. Для демонстрации эффективности предложенного метода обучения была решена практическая задача прогнозирования временного ряда, который описывает среднемесячную температуру в г. Харькове, Украина [26]. Ряд содержал 24 точки и таким образом обучающая выборка содержала 16 точек и 8 точек были в проверочной выборке.

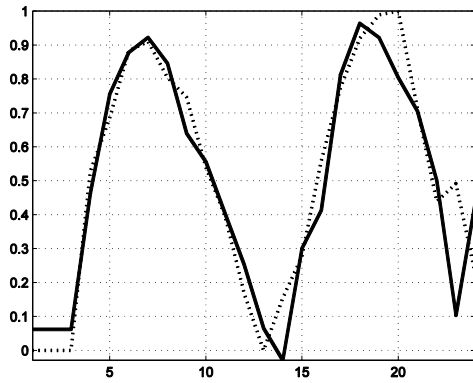
Значения $x(k-2), x(k-1), x(k)$ были взяты для прогнозирования $x(k+1)$. Начальные значения параметра формы вэйвлет-функции активации были взяты $\alpha = 1$. Для оценки качества прогноза была использована среднеквадратичная ошибка (MSE).

На рис. 3 показаны результаты прогнозирования временного ряда с помощью вэйвлет-нейрона с разными методами обучения.

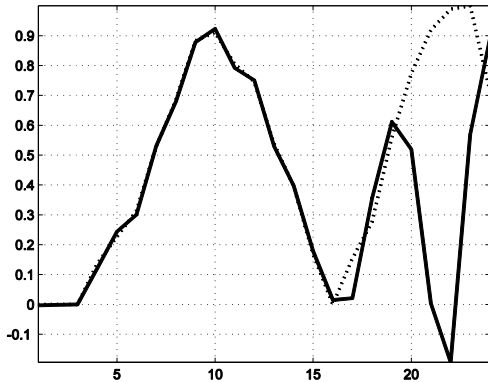
Как видно из рис. 3,а, кривые фактических значений (пунктирная кривая) и прогноза (сплошная кривая) достаточно близки;

на рис. 3,б представлены результаты прогнозирования с помощью вэйвлет-нейрона и градиентного метода обучения с постоянным шагом;

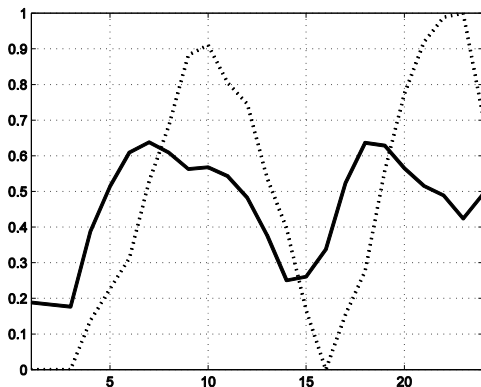
на рис. 3,в показаны результаты вэйвлет-нейрона и предложенного метода обучения (4) и (5).



а)



б)



в)

Рисунок 3 – Результаты прогнозирования временного ряда, описывающего среднемесячную температуру в г. Харьков, Украина:

а) результаты прогноза на основе вэйвлет-нейрона с предложенным методом обучения на основе SVM-критерия;

б) результаты прогнозирования на основе вэйвлет-нейрона с градиентным алгоритмом обучения с постоянным шагом;

в) результаты прогнозирования на основе вэйвлет-нейрона с предложенным алгоритмом обучения всех параметров

В табл. 1 приведены результаты сравнения прогнозирования временного ряда вэйвлет-нейроном с различными методами обучения.

Таблица 1 – Сравнение результатов прогнозирования временного ряда

Нейронная сеть / Метод обучения	Кол. параметров	MSE
Вэйвлет-нейрон / Предложенный метод обучения на основе SVM-критерия	3/8 (on-line)	0,0063
Вэйвлет-нейрон/ Градиентный метод с постоянным шагом	3/8 (10 epoch)	0,000093
Вэйвлет-нейрон/ Предложенным метод обучения всех параметров	3/8 (10 epoch)	0,0374

ВЫВОДЫ. В работе предложена вэйвлет машина опорных векторов наименьших квадратов (LS–WSVM), реализованная на базе вэйвлет-нейрона (WN). Предлагаемая LS–WSVM–WN отличается численной простотой, порождаемой архитектурой WN, небольшим числом настраиваемых параметров, высоким быстродействием, связанным с возможностью применения для обучения процедур оптимизации второго порядка, возможностью обработки последовательно поступающих данных в режиме on-line.

ЛИТЕРАТУРА

1. Du K.-L., Swamy M.N.S. Neural Networks and Statistical Learning. – London: Springer-Verlag, 2014. – 816 p.
2. Haykin S. Neural Networks. A Comprehensive Foundation, Upper Saddle River. – N.J.: Prentice Hall, 1999. – 842 p.
3. Jang J.-S. R. ANFIS: Adaptive-network-based fuzzy inference systems // IEEE Trans. Syst., Man., and Cybern. – 1993. – № 23. – PP. 665–685.
4. Jang J.-S. R., Sun C. T., Mizutani E. Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence. – N.J.: Prentice Hall, 1997. – 640 p.
5. Nelles O. Nonlinear System Identification. – Berlin: Springer, 2001. – 785 p.
6. Uchino E., Yamakawa T. Soft computing based signal prediction, restoration and filtering. Intelligent Hybrid Systems: Fuzzy Logic, Neural Networks and Genetic Algorithms; ed. by Da Ruan. – Boston: Kluwer Academic Publisher, 1997. – PP. 331–349.
7. Bodyanskiy Ye., Vynokurova O. Hybrid adaptive wavelet-neuro-fuzzy system for chaotic time series identification // Information Science. – 2013. – 220. – PP. 170–179.
8. Ljung L. System Identification: Theory for the User. – N.Y.: Prentice-Hall, 1999. – 519 p.
9. Cortes C., Vapnik V. Support vector networks // Machine Learning. – 1995. – № 20. – PP. 273–297.

10. Вапник В.Н., Червоненкіс А.Я. Теория распознавания образов. – М.: Наука, 1974. – 416 с.
11. Вапник В.Н. Восстановление зависимостей по эмпирическим данным. – М.: Наука, 1979. – 448 с.
12. Vapnik V.N. The Nature of Statistical Learning Theory. – N.Y.: Springer, 1995. – 188 p.
13. Suykens J.A.K., Van Gestel T., De Brabanter J., De Moor B., Vandewalle J. Least Squares Support Vector Machines. – Singapore: World Scientific, 2002. – 294 p.
14. Abe S., Tsujinishi D. Fuzzy Least Squares Support Vector Machines for multiclass problems // Neural Networks. – 2003. – № 16. – PP. 785–792.
15. Lin Ch.-F., Wang Sh.-D. Fuzzy Support Vector Machines // IEEE Trans. on Neural Networks. – 2002. – № 13. – PP. 646–671.
16. Pandhiani S.M., Shabri A.B. Time Series Forecasting Using Wavelet-Least Squares Support Vector Machines and Wavelet Regression Models for Monthly Stream Flow Data // Open Journal of Statistics. – 2013. – № 3. – PP. 183–194.
17. Kumar D., Tripathy R.K., Acharya A. Least square support vector machine based Multiclass classification of EEG signals // WSEAS Transactions on Signal Processing. – 2014. – 10. – PP. 86–94.
18. Zahirniak D., Chapman R., Rogers S.K., Suter B.W., Kabrisky M., Pyati V. Pattern recognition using radial basis function network // Application of Artificial Intelligence Conf. Dayton, OH. – 1990. – PP. 249–260.
19. Yamakawa T. A novel nonlinear synapse neuron model guaranteeing a global minimum – Wavelet neuron // Proc. 28th IEEE Int. Symp. On Multiple-Valued Logic, Fukuoka, Japan, IEEE Comp. Soc. – 1998. – PP. 335–336.
20. Bodyanskiy Ye., Vynokurova O., Kharchenko O. Hybrid cascade neural network based on wavelet-neuron // Information Theories and Application. – 2011. – Vol. 18. – 4. – PP. 335–343.
21. Bodyanskiy Ye., Lamonova N., Pliss I., Vynokurova O. An adaptive learning algorithm for a wavelet neural network // Expert Systems. – 2005. – № 22(5). – PP. 235–240.
22. Bodyanskiy Ye., Pliss I.P., Vynokurova O. Flexible neo-fuzzy neuron and neuro-fuzzy network for monitoring of time series properties // Information Technology and Management Science. – 2013. – Vol. 16. – PP. 47–52.
23. Bodyanskiy Ye., Vynokurova O., Yegorova E. Radial-basis-fuzzy-wavelet-neural network with adaptive activation-membership function // Int. J. on Artificial Intelligence and Machine Learning. – 2008. – № 8 (II). – PP. 9–15.
24. Goodwin G.C., Ramadge P.J., Caines R.E. A globally convergent adaptive predictor // Automatica. – 1981. – 17(1). – PP. 135–140.
25. Gantmacher F.R. The Theory of Matrices. – AMS Chelsea Publishing: Reprinted by American Mathematical Society, 2000. – 660 p.
26. Data Market - the open portal to thousands of datasets [[http://datamarket.com/en/data/set/1lo0/#!ds=1lo0!1n6s=2qi.2ql.2qn&display=line&title=Average+monthly+temperatures+across+the+world+\(1701-2011\)&s=8gd](http://datamarket.com/en/data/set/1lo0/#!ds=1lo0!1n6s=2qi.2ql.2qn&display=line&title=Average+monthly+temperatures+across+the+world+(1701-2011)&s=8gd)].

ADAPTIVE LEARNING ALGORITHM FOR WAVELET LEAST SQUARES VECTOR MACHINES

O. Vynokurova, I. Pliss, K. Shkuro. O. Kharchenko

Radio Electronics Kharkiv National University

prosp. Lenina, 14, Kharkiv, 61166, Ukraine. E-mail: vinokurova@kture.kharkov.ua, pliss@kture.kharkov.ua

Special type of neural networks, whose learning is based on both optimization and memory, are support vector machines (SVM). Their architecture aligns with radial basis function networks and general regression neural networks, and synaptic weights are defined as a result of non-linear programming tasks solving, and centers of activation functions are defined on the principle “neurons in data points”. And although SVM-networks have many definite advantages, their learning process (from computational complexity point of view) is sufficiently time-consuming because of solving of non-linear programming tasks of large dimension. In connection with this the least squares support vector machines (LS-SVM) were proposed as alternative of SVM. Learning process in these networks is reduced to the solving of linear equation systems. It is much simpler from computational complexity point of view and such learning process can be implemented in on-line mode. Wavelet-modification of conventional SVM is so-called wavelet support vector machines (WSVM), in which multidimensional kernel activation functions are replaced by one-dimensional wavelet-functions. And although WSVM have large possibilities in comparison with conventional SVM, their learning process is connected with realization of quite complex procedures, what limits their ability for solving tasks in on-line mode. Therefore it seems like a good idea to develop simple wavelet-neuro-system that realizes learning ideas based on minimization of empirical risk and oriented on information processing in on-line mode. As elementary block of such systems we propose to use wavelet-neuron that has improved approximation properties, computational simplicity, high learning rate

and capability of hidden features identification in data processing. The architecture and learning algorithm for least squares wavelet support machines, which are characterized by high speed of operation and possibility of learning under conditions of short training set, are proposed.

Key words: Least squares support vector machine, wavelet-neuron, forecasting, non-linear non-stationary time series, adaptive wavelet function.

REFERENCES

1. Du, K.-L., Swamy, M.N.S. (2014) *Neural Networks and Statistical Learning*, London, Springer-Verlag.
2. Haykin, S. (1999) *Neural Networks. A Comprehensive Foundation* Upper Saddle River, N.J., Prentice Hall.
3. Jang, J.-S. R. (1993) "ANFIS: Adaptive-network-based fuzzy inference systems," *IEEE Trans. on Syst. Man. and Cybern.*, no. 23, pp. 665–685.
4. Jang, J.-S. R., Sun, C. T., Mizutani, E. (1997) *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, N.J., Prentice Hall.
5. Nelles, O. (2001) *Nonlinear System Identification*, Berlin, Springer.
6. Uchino, E., Yamakawa, T. (1997) "Soft computing based signal prediction, restoration and filtering," in *Intelligent Hybrid Systems: Fuzzy Logic, Neural Networks and Genetic Algorithms*, Da Ruan Eds., Boston: Kluwer Academic Publisher, pp. 331–349.
7. Bodyanskiy, Ye., Vynokurova, O. (2013) "Hybrid adaptive wavelet-neuro-fuzzy system for chaotic time series identification," *Information Science*, no. 220, pp. 170–179.
8. Ljung, L. (1999) *System Identification: Theory for the User*, N.Y., Prentice-Hall.
9. Cortes, C., Vapnik, V. (1995) "Support vector networks," *Machine Learning*, n. 20, pp. 273–297.
10. Vapnik, V.N., Chervonenkis, A.Ya. (1974) *Pattern Recognition Theory (Statistical Learning Problems)*, Nauka, Moscow, Russia.
11. Vapnik, V.N. (1979) *Empirical Data Dependencies Restoration*, Nauka, Moscow, Russia.
12. Vapnik, V.N. (1995) *The Nature of Statistical Learning Theory*, N.Y., Springer.
13. Suykens, J.A.K., Van Gestel, T., De Brabanter, J., De Moor B., Vandewalle, J. (2002) *Least Squares Support Vector Machines*, Singapore: World Scientific.
14. Abe, S., Tsujinishi, D. (2003) "Fuzzy Least Squares Support Vector Machines for multiclass problems," *Neural Networks*, no. 16, pp. 785–792.
15. Lin, Ch.-F., Wang, Sh.-D. (2002) "Fuzzy Support Vector Machines," *IEEE Trans. on Neural Networks*, no. 13, pp. 646–671.
16. Pandhiani, S.M., Shabri, A.B. (2013) "Time Series Forecasting Using Wavelet-Least Squares Support Vector Machines and Wavelet Regression Models for Monthly Stream Flow Data," *Open Journal of Statistics*, no. 3, pp. 183–194.
17. Kumar, D., Tripathy, R.K., Acharya, A. (2014) "Least squares support vector machine based multiclass classification of EEG signals," *WSEAS Transactions on Signal Processing*, vol. 10, pp. 86–94.
18. Zahirniak, D., Chapman, R., Rogers, S.K., Suter, B.W., Kabrisky, M., Pyati, V. (1990) "Pattern recognition using radial basis function network," in *Application of Artificial Intelligence Conf., Dayton, OH*, pp. 249–260.
19. Yamakawa, T. (1998) "A novel nonlinear synapse neuron model guaranteeing a global minimum – Wavelet neuron," in *Proc. 28th IEEE Int. Symp. On Multiple-Valued Logic, Fukuoka, Japan, IEEE Comp. Soc.*, pp. 335–336.
20. Bodyanskiy, Ye., Vynokurova, O., Kharchenko, O. (2011) "Hybrid cascade neural network based on wavelet-neuron," *Information Theories and Application*, vol. 18, no. 4, pp. 335–343.
21. Bodyanskiy, Ye., Lamonova, N., Pliss, I., Vynokurova, O. (2005) "An adaptive learning algorithm for a wavelet neural network," *Expert Systems*, vol. 22, no. 5, pp. 235–240.
22. Bodyanskiy, Ye., Pliss, I., Vynokurova, O. (2013) "Flexible neo-fuzzy neuron and neuro-fuzzy network for monitoring of time series properties," *Scientific J. of Riga Technical University. Information Technology and Management Science*, vol.16, pp. 47–52.
23. Bodyanskiy, Ye., Vynokurova, O., Yegorova, E. (2008) "Radial-basis-fuzzy-wavelet-neural network with adaptive activation-membership function" *Int. J. on Artificial Intelligence and Machine Learning*, no. 8 (II), pp. 9-15.
24. Goodwin, G.C., Ramadge, P.J., Caines, R.E. (1981) "A globally convergent adaptive predictor," *Automatica*, vol. 17, no. 1, pp. 135–140.
25. Gantmacher, F.R. (2000) *The Theory of Matrices* AMS Chelsea Publishing: Reprinted by American Mathematical Society.
26. Data Market - the open portal to thousands of datasets [[http://datamarket.com/en/data/set/1loo#!ds=1loo!In6s=2qi.2ql.2qn&display=line&title=Average+monthly+temperatures+across+the+world+\(1701-2011\)&s=8gd](http://datamarket.com/en/data/set/1loo#!ds=1loo!In6s=2qi.2ql.2qn&display=line&title=Average+monthly+temperatures+across+the+world+(1701-2011)&s=8gd)].

Стаття надійшла 24.09.14.