УДК 004.852

О.Г. Руденко[1], О.О. Безсонов[2], О. С. Романюк [3]

[1] ХНЕУ, м. Харків, Україна, oleg.rudenko@hneu.net;
[2] ХНЕУ, м. Харків, Україна, oleksandr.bezsonov@hneu.net;
[3] ХНУРЕ, м. Харків, Україна, romanyk@gmail.com

ПРО ОДИН АЛГОРИТМ НАВЧАННЯ НЕЙРОННОЇ МЕРЕЖІ В ЗАДАЧІ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

У статті запропоновано метод навчання нейронних мереж при вирішенні задачі прогнозування часового ряду (ЧР). Більшість практичних задач прогнозованя ЧР характеризуються високим рівнем нелінійності і нестаціонарності, зашумленістю, наявністю нерегулярних трендів, стрибків, аномальних викидів. У цих умовах жорсткі статистичні припущення про властивості ЧР часто обмежують можливості класичних методів прогнозування. Альтернативою статистичним методам можуть служити методи обчислювального інтелекту, до числа яких відносяться штучні нейронні мережі. Результати імітаційного моделювання підтвердили, що запропонований метод навчання нейронної мережі дозволяє значно підвищити точність прогнозування часових рядів.

ЧАСОВИЙ РЯД, ШТУЧНА НЕЙРОННА МЕРЕЖА, АЛГОРИТМ НАВЧАННЯ, АКТИВАЦІЙНА ФУНКЦІЯ, ІМІТАЦІЙНЕ МОДЕЛЮВАННЯ

Руденко О.Г., Бессонов А.А., Романюк А.С. Об одном алгоритме обучения нейронных сетей в задачах прогнозирования временных рядов. В статье предложен метод обучения нейронных сетей при решении задачи прогнозирования временного ряда (ВР). Большинство практических задач прогнозирования ВР характеризуются высоким уровнем нелинейности и нестационарности, зашумленности, наличием нерегулярных трендов, скачков, аномальных выбросов. В этих условиях жесткие статистические предположения о свойствах ВР часто ограничивают возможности классических методов прогнозирования. Альтернативой статистическим методам могут служить методы вычислительного интеллекта, к числу которых относятся искусственные нейронные сети. Результаты имитационного моделирования подтвердили, что предложенный метод обучения нейронной сети позволяет значительно повысить точность прогнозирования временных рядов.

ВРЕМЕННОЙ РЯД, ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ, АЛГОРИТМОВ ОБУЧЕНИЯ, АКТИВАЦИОННАЯ ФУНКЦИЯ, ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ

Rudenko O.G., Bezsonov O.O., Romanyk O.S. About one algorithm of training the neural network in the problem of the time rseries prediction. The article proposes a method of neural networks training in solving the problem of prediction of the time series. Most of the predictive tasks of the time series are characterized by high levels of nonlinearity and non-stationary, noisiness, irregular trends, jumps, abnormal emissions. In these conditions, rigid statistical assumptions about the properties of the time series often limit the possibilities of classical forecasting methods. The alternative methods to statistical methods can be the methods of computational intelligence, which include artificial neural networks. The simulation results confirmed that the proposed method of training the neural network can significantly improve the prediction accuracy of the time series.

TIME SERIES, ARTIFICIAL NEURAL NETWORK, TRAINING ALGORITHM, ACTIVATION FUNCTION, SIMULATION

Вступ

Задача прогнозування часових рядів (ЧР) була і залишається актуальною, особливо останнім часом, коли стали доступні потужні засоби збору та обробки інформації. Прогнозування часових рядів є важливою науково-технічною проблемою, так як дозволяє передбачити поведінку різних факторів в екологічних, економічних, соціальних та інших системах.

Розвиток прогностики як науки в останні десятиліття призвело до створення безлічі моделей і методів, процедур, прийомів прогнозування, нерівноцінних за своїм значенням. За оцінками зарубіжних і вітчизняних фахівців з прогностиці вже налічується понад ста методів прогнозування, в зв'язку з чим постає завдання вибору методів, які давали б адекватні прогнози для досліджуваних процесів або систем.

До останнього часу основним при вирішенні задачі прогнозування був статистичний підхід. В рамках статистичних моделей вирішуються задачі прогнозування, знаходження прихованих періодичностей в даних, аналізу залежностей, оцінки ризиків при прийнятті рішень та інші. Загальним недоліком статистичних моделей є складність вибору типу моделі і підбору її параметрів. Крім того, при використанні статистичного підходу одним з головних вимог до часового ряду є його стаціонарність, яка полягає в тому, що розподіл його значень є інваріантним щодо моменту часу, для якого воно побудовано.

Традиційні статистичні методи аналізу ЧР часто виявляються неефективними з огляду на те, що ці методи потребують апріорної наявності досить великої і репрезентативної вибірки.

Слід, однак, відзначити, що більшість практичних задач прогнозовані ЧР характеризуються високим рівнем нелінійності і нестаціонарності, зашумленістю, наявністю нерегулярних трендів, стрибків, аномальних викидів. У цих умовах жорсткі статистичні припущення про властивості ЧР часто обмежують можливості класичних методів прогнозування [1].

Альтернативою статистичним методам можуть служити методи обчислювального інтелекту, до числа яких, в першу чергу, слід віднести штучні нейронні мережі (ШНМ) [1-6]. Будучи універсальними апроксиматорами, деякі типи ШНМ дозволяють відновити з заданою точністю будь-яку як завгодно складну безперервну нелінійну функцію, використовуючи уявлення функції, що апроксимується у вигляді нейронної мережі, утвореної нейронами, параметри яких визначаються шляхом її навчання [7-9]. Здатність нейронної мережі до різнобічної обробці інформації випливає з її здатності до узагальнення і виділення прихованих залежностей між вхідними та вихідними даними. Істотною перевагою нейронних мереж є те, що вони здатні до навчання і узагальнення накопичених знань.

У даній роботі розглядається задача прогнозування часового ряду засобами нейронних мереж.

1. Багатошаровий персептрон як нейромережева модель

Метою будь-якого прогнозування є створення моделі, яка дозволяє прогнозувати майбутнє і оцінити тенденції в змінах того чи іншого фактора. Якість прогнозу в такому випадку залежить від наявності передісторії змінюваного чинника, похибок вимірювання даної величини і інших чинників.

Побудова математичної моделі, використовуваної при прогнозуванні, пов'язана з апроксимацією деяких, в загальному випадку нелінійних, функцій

$$y(x) = f(x) + \xi\,, \tag{1}$$

де $x = \left[x_1, x_2 \ldots, x_N\right]^T$ — вектор узагальненого вхідного сигналу $\mathrm{N} \times 1$; $f(\bullet)$ — невідома нелінійна функція; ξ — завада; T — символ транспонування, тобто полягає в отриманні оцінки функції $f(\bullet)$ за вимірюваннями вхідних і вихідних змінних.

Відсутність інформації про вид нелінійності часто призводить до неефективності традиційних методів апроксимації, а в ряді випадків — до їх непридатності.

З іншого боку, деякі типи ШНМ дозволяють відновити з заданою точністю будь-яку як завгодно складну безперервну нелінійну функцію, використовуючи уявлення функції, що апроксимується у вигляді нейронної мережі, утвореної нейронами, параметри яких визначаються шляхом її навчання на основі пред'явлення навчаючих пар $\{\mathrm{x}(k), y(k)\}, k = 1, 2, \ldots$.

Серед найбільш широко застосовуваних для вирішення цієї задачі ШНМ (багатошаровий персептрон (БП), радіально-базисні мережі (РБМ) і мережа CMAC (Cerebellar Model Articulation Controller) [3–6]), досить ефективним є БП, який використовує апроксимації нелінійного оператора $f(\bullet)$ виду

$$\hat{y}(x) = \hat{f}(x) =$$
$$= f^q\left[\left(W^q\right)^T f^{q-1}\left[\left(W^{q-1}\right)^T f^{q-2}\left[\ldots f^1\left[\left(W^1 x + b_1\right)^T\right]\ldots\right]\right]\right] +$$
$$+ b_q, \tag{2}$$

де W^i — вектор вагових параметрів нейронів i-го шару мережі; $f^i[\bullet]$ — активаційна функція (АФ) i-го шару; b_i — зміщення i-го нейрона.

В якості АФ найчастіше використовуються такі функції

$$f_{th}(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{3}$$

та

$$f_{\log}(x) = \frac{1}{1 + e^x}. \tag{4}$$

Існують два основних підходи для надання нейронним мережам на базі БП властивостей, необхідних для обробки динамічних даних: додавання елементів затримок сигналу на вході мережі та додавання рекурентних зв'язків до внутрішньої структури мережі. В першому випадку, при використанні «методу часового вікна» [3-6], нейронна мережа отримує на вхід разом з поточним вхідни затримані в часі минулі значення вхідних сигналів. Навчання нейромережі здійснюється за відомим методом зворотного поширення (Backpropagation, BP) і з застосуванням градієнтних методів оптимізації (зокрема, методів Гауса-Ньютона, Левенберга-Маркуардта тощо). Незважаючи на простоту й технологічність цього підходу (за оцінками [10], зараз більш ніж в 90% випадків для прогнозування часових рядів використовується саме така схема), при його застосуванні необхідно апріорі визначити розмір часового вікна (це відповідає визначенню структури мережі), що суттєво впливає на якість прогнозування . Слід зазначити, що при використанні структури моделі не адекватній структурі динамічного процесу, залишається відкритим питання навчання нейромережі. Крім того, якщо при такому підході для багатокрокового прогнозу в якості вхідних даних використовуються власні прогнозні дані нейромережі, отримані для попередніх кроків, якість такого прогнозу буде низькою.

Іншим підходом введення динаміки в нейромережі прямого поширення є додавання внутрішніх рекурентних зв'язків у приховані, вхідні або вихідні шари нейромережі [2-4,10,12]. При цьому для визначення впливу минулих тактів на поточний результат вводиться обчислення спеціальних динамічних похідних шляхом «рекурентного навчання в реальному часі» або «зворотним поширенням в часі». Налаштування ваг нейромережі здійснюється також будь-яким градієнтним алгоритмом. Такі рекурентні мережі за своєю структурою більш відповідають динамічним процесам, що моделюються, і тому краще показують себе в задачах керування і багатокрокового прогнозування. Разом з тим навчання таких мереж є більш важким завданням через додаткові ступені свободи в таких мережах. Крім того, при синтезі таких мереж необхідно досліджувати питання їх сталості, а при розрахунках динамічних похідних у персептроноподібних мережах має місце ефект зникнення градієнта [3].

Таким чином, використання ШНМ вимагає вирішення задач структурної та параметричної оптимізації, відповідних вибору оптимальної топології мережі і її навчання (налаштування параметрів — в БП це вагові параметри і параметри АФ).

2. Алгоритм навчання мережі при відсутності завад $\xi(k)$

Розглянемо задачу прогнозування часового ряду, що описується рівнянням

$$y(k) = f(z(k)) = f(w^{*T} x(k)), \tag{5}$$

де $z(k) = w^{*T} x(k), x(k) = (x_1(k), x_2(k), \ldots x_N(k))^T$ – вектор вхідних сигналів $N \times 1$;
$x_i(k) = x(k-i-1), i = 1,2,\ldots N$; $f(\bullet)$ — невідома нелінійна функція; $w^* = (w^*_1(k), w^*_2(k), \ldots w^*_N(k))^T$ – вектор вагових параметрів $N \times 1$; T — символ транспонування.

Помилка прогнозування визначається таким чином

$$e(k) = y^*(k) - y(k), \tag{6}$$

де $y^*(k)$ – сигнал, що прогнозується.

Точність прогнозування будемо оцінювати за критерієм

$$E(k) = 0{,}5e^2(k). \tag{7}$$

Як вже зазначалося вище, задача навчання ШНМ полягає в налаштуванні вектора її параметрів w(k) за формулою

$$w(k+1) = w(k) + \Delta w(k), \tag{8}$$

де

$$\Delta w(k) = -\gamma \nabla_w [E(k)]_{w=w(k)} \tag{9}$$

– приріст значень вектора параметрів; γ – коефіцієнт, що впливає на час навчання.

Для отримання правила налаштування розглянемо розклад помилки прогнозування на k+1-у кроці в ряд Тейлора і обмежемось лінійними членами

$$e(k+1) = e(k) + \left(\frac{\partial e(k)}{\partial w(k)}\right)^T \Delta w(k) + \left(\frac{\partial e(k)}{\partial x(k)}\right)^T \Delta x(k) + \frac{\partial e(k)}{\partial y^*(k)} \Delta y^*(k) + \ldots. \tag{10}$$

У зв'язку з цим квадрат помилки можна представити таким чином

$$e^2(k+1) = e^2(k) + 2e(k)\left(\frac{\partial e(k)}{\partial w(k)}\right)^T \Delta w(k) + \left\| \left(\frac{\partial e(k)}{\partial w(k)}\right)^T \Delta w(k) \right\|^2 \tag{11}$$

Враховуючи (5), (6) можна отримати наступні вирази для похідних в рівнянні (11)

$$\frac{\partial e(k)}{\partial w(k)} = -f'(z(k)) x^T(k), \tag{12}$$

а

$$\left\| \left(\frac{\partial e(k)}{\partial w(k)}\right)^T \Delta w(k) \right\|^2 = \Delta w^T(k) \frac{\partial e(k)}{\partial w(k)} \left(\frac{\partial e(k)}{\partial w(k)}\right)^T \Delta w(k).$$

Це дозволяє записати (9) наступним чином

$$\Delta w(k) = -\gamma e(k) f'(z(k)) x(k). \tag{13}$$

Врахувавши, що

$$\Delta w^T(k) \frac{\partial e(k)}{\partial w(k)} \left(\frac{\partial e(k)}{\partial w(k)}\right)^T \Delta w(k) = \Delta w^T(k)(-f'(z(k))x(k))(x^T(k)(-f'(z(k))))\Delta w(k),$$

можна записати наступний вираз для $e(k+1)$:

$$e(k+1) \approx e(k)[1 - \gamma \left| f'(z(k)\right|^2 \|x(k)\|^2],$$

що дозволяє представити $e^2(k+1)$ так

$$e^2(k+1) = e^2(k)\left[1 - 2\gamma (f'(z(k)))^2 \|x(k)\|^2 + \gamma^2 (f'(z(k)))^4 \|x(k)\|^4\right].$$

Для збіжності алгоритму (8, 9) необхідне виконання нерівності

$$\left| 1 - 2\gamma (f'(z(k)))^2 \|x(k)\|^2 + \gamma^2 (f'(z(k)))^4 \|x(k)\|^4 \right| < 1,$$

звідки маємо

$$0 < \gamma < \frac{2}{(f'(z(k)))^2 \|x(k)\|^2}. \tag{14}$$

З умови

$$\frac{\partial e^2(k+1)}{\partial \gamma} = 0$$

можна отримати вираз для оптимального значення параметра γ, при якому швидкість збіжності алгоритма навчання буде максимальною

$$\gamma^{\text{опт}}(k) = \frac{1}{(f'(z(k)))^2 \|x(k)\|^2}, \tag{15}$$

тобто отримуємо алгоритм Качмажа (Уідроу-Хоффа) [3-6].

При застосуванні АФ виду (3), (4) похідні, що використовуються в (14), (15) легко обчислюються за формулами (для нашого випадку)

$$f'_{th}(z(k)) = z(k)(1 - f^2_{th}(z(k))$$

$$f'_{\log}(z(k)) = z(k) f_{\log}(z(k))(1 - f_{\log}(z(k))$$

або

$$f'_{th}(w^T x) = x(k)(1 - f^2_{th}(w^T x)),$$

$$f'_{\log}(w^T x) = x(k) f_{\log}(z(k))(1 - f_{\log}(w^T x).$$

Для підвищення обчислювальної стійкості процедури навчання слід скористатися регуляризацією алгоритму навчання [16-18], тобто замість (15) взяти

$$\gamma'(k) = \frac{1}{(f'(z(k)))^2 \|x(k)\|^2 + \delta(k)}, \qquad (16)$$

де $\delta(k) > 0$ – параметр регуляризації.

По аналогії з (14) неважко отримати, що для збіжності регуляризованого алгоритму необхідне виконання нерівності

$$0 < \frac{1}{(f'(z(k)))^2 \|x(k)\|^2 + \delta(k)} < \frac{2}{(f'(z(k)))^2 \|x(k)\|^2}. \qquad (17)$$

З нерівності (17) отримуємо, що регуляризований алгоритм буде збігатися , якщо параметр параметр регуляризації $\delta(k)$ задовольняє нерівності

$$\delta(k) > -\frac{1}{2}(f'(z(k)))^2 \|x(k)\|^2. \qquad (18)$$

Як видно з (18), параметр $\delta(k)$ є змінним у часі і може коригуватися на кожному кроці при надходженні нової інформації, наприклад, таким чином

$$\delta(k) = \delta(k-1) - \alpha \nabla_{\delta} [E(k)]_{\delta=\delta(k)}. \qquad (19)$$

Така процедура була розглянута в роботі [17]. Однак при цьому, на що не звернули увагу автори цієї роботи, виникає задача визначення оптимального значення параметра α. Тому доцільним є використання параметра $\delta(k) = \delta = const$, величина якого визначається окремо для кожної конкретної задачі. Деякі рекомендації щодо вибору δ можна знайти в [16–18].

3. Навчання мережі при наявності обмежених завад $\xi(k)$

Якщо щодо перешкоди відомо тільки, що вона обмежена по амплітуді

$$|\xi(k)| < \beta, \qquad (20)$$

для оцінювання параметрів застосовуються методи, які не використовують ніякої інформації про статистичні властивості збурень крім їх приналежності до деякого обмеженого інтервалу.

В даний час для оцінювання параметрів при наявності обмежених за амплітудою завад найбільш широкого поширення набули алгоритми, в основі побудови яких лежить метод політопів (і як окремий випадок — ортотопів), алгоритми, що містять зону нечутливості і алгоритми на основі побудови еліпсоїдів.

Якщо про перешкоди відомо, що вони задовольняють умові (20), то ця інформація може бути врахована в алгоритмі, що містить зону нечутливості. При цьому знижуються вимоги щодо знання властивостей завади, однак огрублюється і сам алгоритм навчання.

Ідея використання в алгоритмі зони нечутливості заснована на тому, що навіть в разі точного визначення параметрів моделі залишається помилка (неузгодженість між вихідними сигналами об'єкта і моделі), величина якої визначається величиною завади (20).

Широке поширення набуло використання елементів, що реалізують зону нечутливості, яка описується співвідношенням

$$g(e(k),\beta) = \begin{cases} e(k) - \beta(k) signe(k), & \text{якщо} \quad |e(k)| \geq \beta; \\ 0, & \text{якщо} \quad |e(k)| < \beta. \end{cases} \qquad (21)$$

Розглянемо градієнтний алгоритм виду

$$w(k+1) = w(k) + \gamma g(e(k),\beta) f^{'}(z(k)) x(k), \qquad (22)$$

що містить зону нечутливості $g(e(k),\beta)$.

Тут $\gamma > 0$ — деякий параметр, що впливає на швидкість збіжності.

Введемо помилку оцінювання

$$\Theta(i) = w^{*} - w(i), \; i = 1,2,...$$

і функцію Ляпунова виду $V(i) = \left\| \Theta(i) \right\|^2$.

Після вирахування з обох частин (22) w^{*} маємо

$$\Theta(k+1) = \Theta(k) - \gamma g(e(k),\beta) f^{'}(z(k)) x(k). \qquad (23)$$

Помноживши зліва обидві частини (23) на $\Theta^T(k+1)$, отримуємо

$$\begin{aligned} \|\Theta(k+1)\|^2 &= \|\Theta(k)\|^2 - \\ &-2\gamma g(e(k),\beta) f^{'}(z(k)) \Theta^T(k) x(k) + \\ &+\gamma^2 g^2(e(k),\beta)(f^{'}(z(k)))^2 \|x(k)\|^2. \end{aligned} \qquad (24)$$

Розглянемо приріст функції Ляпунова

$$\Delta V(k+1) = V(k+1) - V(k). \qquad (25)$$

Для збіжності алгоритму необхідно, щоб

$$\Delta V(k+1) < 0. \qquad (26)$$

Підстановка (24) в (25) дає

$$\Delta V(k+1) = -2\gamma g(e(k),\beta) f^{'}(z(k)) \Theta^T(k) x(k) + \\ +\gamma^2 g^2(e(k),\beta)(f^{'}(z(k)))^2 \|x(k)\|^2.$$

Приймаючи до уваги, що

$$\Theta^T(k) x(k) f^{'}(z(k)) = e(k) - \xi(k) \;\; \text{і} \;\; \gamma > 0,$$

можна визначити, що для виконання (26), необхідно, щоб

$$g^2(e(k),\beta) \leq \frac{2(e(k) - \xi(k) g(e(k),\beta)}{\gamma (f^{'}(z(k)))^2 \|x(k)\|^2}. \qquad (27)$$

У зв'язку з тим, що

$$g^2(e(k),\beta) = |g(e(k),\beta)| signg(e(k),\beta),$$

а

$$signg(e(k),\beta) = \mathrm{sign}\, e(k),$$

умова збіжності (27) буде мати вигляд

$$|g(e(k),\beta)| \leq \frac{2}{\gamma (f^{'}(z(k)))^2 \|x(k)\|^2} [|e(k)| - \xi(k)] signe(k).$$

Крім того, з огляду на те, що завада $\xi(k)$ обмежена (22) і навчання відбувається при $|e(k)| \geq \delta$, дану умову можна переписати так:

$$|g(e(k),\beta)| \le \frac{2}{\gamma(f^{'}(z(k)))^2\|x(k)\|^2}[|e(k)|-\xi(k)]. \quad (28)$$

Таким чином, значення $g(e(k),\beta)$ залежить як від величини вільно обраного параметра γ, так і від значень $f^{'}(z(k)))^2\|x(k)\|^2$.

Аналогічно для алгоритму

$$w(k+1) = w(k) + \gamma\frac{g(e(k),\delta)f^{'}(z(k))x(k)}{(f^{'}(z(k)))^2\|x(k)\|^2}, \quad (29)$$

де $\gamma > 0$ — деякий параметр, неважко показати, що його збіжність буде спостерігатися при виконанні умови

$$|g(e(k),\beta)| \le \frac{2}{\gamma}[|e(k)-\delta|]. \quad (30)$$

Для регуляризоване алгоритму

$$w(k+1) = w(k) + \gamma\frac{g(e(k),\delta)f^{'}(z(k))x(k)}{(f^{'}(z(k)))^2\|x(k)\|^2+\delta} \quad (31)$$

нескладно отримати таку умову збіжності:

$$|g(e(k),\beta)| \le \frac{2A}{\gamma}(|e(k)-\beta)|, \quad (32)$$

де $A = 1+\frac{\delta}{(f^{'}(z(k)))^2\|x(k)\|^2}$.

Як видно з наведених результатів, для монотонної збіжності процедур повинні виконуватися відповідні умови на кожному такті процесу навчання. Так як ці умови збіжності накладають цілком певні обмеження на величину зони нечутливості, доцільно ввести її корекцію зон на кожному такті. З іншого боку, параметри зон нечутливості залежать від характеристик присутніх перешкод. Якщо наявність інформації про характеристики перешкоди (зокрема, значення параметра β, що входить в (22)) спрощує вибір зони нечутливості, то її відсутність суттєво ускладнює її вибір. Тому доцільним видається адаптивна настройка величини зони нечутливості, здійснювана в міру надходження нової інформації.

Висновки

В роботі наведено результати дослідження властивостей градієнтного алгоритму навчання нейромережі (багатошарового персептрону). Отримано умови збіжності алгоритму і вираз для оптимального значення параметру збіжності, яке забезпечує максимальну швидкість навчання. Крім того, визначено умови збіжності для регуляризованого алгоритму навчання. Показано, що можливим є визначення значення оптимального параметру регуляризації, який є змінним у часі, шляхом мінімізації деякого, зокрема, квадратичного функціоналу. Однак при цьому виникає питання визначення оптимального іншого параметра. Тому доцільним є використання постійного параметра регуляризації, значення якого залежать від необхідної точності задачі, що вирішується.

Отримані умови збіжності алгоритмів навчання, які використовують зону нечутливості, за наявністю обмежень завад. Показано, що параметри цієї зони залежався від характеристик присутніх завад. Відсутність такої інформації суттєво ускладнює вибір величини зони нечутливості и обумовлює доцільність її адаптивного налаштування, яке здійснюється при надходженні нової інформації.

Список літератури:

1. *Снитюк В. Е.* Прогнозирование. Модели, методы, алгоритмы: учебное пособие. / В.Е. Снитюк / – К.: «Маклаут», 2008. – 364 с. **2.** *Mandic D. P.* Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability. / D. P. Mandic, J. A. Chambers / John Wiley & Sons, 2001. – 285 p. https://doc.lagout.org/science/0_ComputerScience/3_Theory/NeuralNetworks/RecurrentNeuralNetworksforPrediction.pdf **3.** *Хайкин С.* Нейронные сети: полный курс / С. Хайкин / Пер. с англ. – М.: Вильямс, 2006. – 1104 с. **4.** *Осовский С.* Нейронные сети для обработки информации / С. Осовский / Пер. с польского. – М.: Финансы и статистика, 2002. – 344 с. **5.** *Руденко О.Г.* Основы теории искусственных нейронных сетей / О.Г. Руденко, Е.В. Бодянский. – Харьков: ТЕЛЕТЕХ, 2002. – 317 с. **6.** *Бодянский Е.В.* Искусственные нейронные сети: архитектура, обучение, применение / Е.В. Бодянский, О.Г. Руденко. – Харьков: ТЕЛЕТЕХ, 2004. – 372 с. **7.** *Cybenko G.* Approximation by superposition of a sigmoidal function / G. Cybenko / Math. of controls, Signals & Systems. – 1989. – 2. – P.303-314. **8.** *Hornik K.* Multilayer feedforward networks are universal approximators/ K. Hornik, M. Stinchcombe, H. White / Neural Networks. – 1989. – v.2. – P.359-366. **9.** *Poggio T.* Networks for approximation and learning / T. Poggio, F. Giorosi / Proc. IEEE. – 1990. – V.78. — №9. – P. 1481-1497. **10.** *Чернодуб А. М.* Навчання рекурентних нейронних мереж методом псевдорегуляризації для багатокрокового прогнозування часових рядів / А. М. Чернодуб / Математичні машини і системи. – 2012. – No 4. – С. 41–51. **11.** *Parras-Gutierrez E.* Coevolution of lags and RBFNs for time series forecasting: L-Co-R algorithm. / E. Parras-Gutierrez, M. Garcia-Arenas,V.M. Rivas, V.J. Jesus / Soft Comput / 2012. -16(6). – P. 919–942. **12.** *Assaad M.* A new boosting algorithm for improved time-series forecasting with recurrent neural networks /M. Assaad, R. Boner, H. Cardot / Information Fusion. –2008. –9. – P. 41–55. **13.** *Samsudin R.* A comparison of time series forecasting using support vector machine and artificial neural network model. / R. Samsudin, F. Shabri, P. Saad / Journal of Applied Sciences.-2010. — 10(11). — P. 950–958. **14.** *Samanta B.* Prediction of chaotic time series using computational intelligence./ B. Samanta / Expert Syst Appl. –2011. –38(9). – P.11406–11411. **15.** *Maria J.* Long-term time series prediction with the NARX network: an empirical evaluation. / J. Maria, J.P. Menezes, G.A. Barreto / Neurocomputing. – 2008. –71(16-18). – P. 3335-43. **16.** *Тихонов А.Н.* Методы решения некорректных задач. / А.Н. Тихонов, В.Я. Арсенин / М.: Наука, 1980. - 223 с. **17.** *Демиденко Е.З.* Линейная и нелинейная регрессии./ Е.З. Демиденко /– М.: Финансы и статистика, 1981. - 302 с. **18.** *Райбман Н.С.* Адаптивные модели в системах управления / Н.С. Райбман, В.М. Чадеев /– М.: Сов. радио, 1966. –159 с. **19.** *Goh S.L.* A Nonlinear Neural FIR Filter With An Adaptive Activation Function/ S.L. Goh, M. Bozic, D. P. Mandic / Journal of automatic control, university of Belgrade. – 2003. – Vol. 13(1). – P. 1–5.

Надійшла до редколегії 26.01.2018