“Neural network time series prediction based on multilayer perceptron”

AUTHORS	Oleg Rudenko https://orcid.org/0000-0003-0859-2015 Oleksandr Bezsonov https://orcid.org/0000-0001-6104-4275 Oleksandr Romanyk https://orcid.org/0000-0003-3278-1772

ARTICLE INFO	Oleg Rudenko, Oleksandr Bezsonov and Oleksandr Romanyk (2019). Neural network time series prediction based on multilayer perceptron . *Development Management*, *17*(1), 23-34. doi:10.21511/dm.5(1).2019.03
DOI	http://dx.doi.org/10.21511/dm.5(1).2019.03
RELEASED ON	Tuesday, 07 May 2019
RECEIVED ON	Saturday, 01 December 2018
ACCEPTED ON	Thursday, 20 December 2018

LICENSE	This work is licensed under a Creative Commons Attribution 4.0 International License

JOURNAL	"Development Management"
ISSN PRINT	2413-9610
ISSN ONLINE	2663-2365
FOUNDER	Simon Kuznets Kharkiv National University of Economics

NUMBER OF REFERENCES	NUMBER OF FIGURES	NUMBER OF TABLES
41	1	0

© The author(s) 2021. This publication is an open access article.

Oleg Rudenko (Ukraine), **Oleksandr Bezsonov** (Ukraine),
Oleksandr Romanyk (Ukraine)

NEURAL NETWORK TIME SERIES PREDICTION BASED ON MULTILAYER PERCEPTRON

Abstract

Until recently, the statistical approach was the main technique in solving the prediction problem. In the framework of static models, the tasks of forecasting, the identification of hidden periodicity in data, analysis of dependencies, risk assessment in decision making, and others are solved. The general disadvantage of statistical models is the complexity of choosing the type of the model and selecting its parameters. Computing intelligence methods, among which artificial neural networks should be considered at first, can serve as alternative to statistical methods. The ability of the neural network to comprehensively process information follows from their ability to generalize and isolate hidden dependencies between input and output data. Significant advantage of neural networks is that they are capable of learning and generalizing the accumulated knowledge.
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.

Keywords time series, artificial neural network, training algorithm, activation function, simulation

JEL Classification C45

О.Г. Руденко (Україна), **О.О. Безсонов** (Україна),
О.С. Романюк (Україна)

НЕЙРОМЕРЕЖЕВЕ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ НА ОСНОВІ БАГАТОШАРОВОГО ПЕРСЕПТРОНА

Анотація

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. Альтернативою статистичним методам можуть служити методи обчислювального інтелекту, до числа яких, в першу чергу, слід віднести штучні нейронні мережі. Здатність нейронної мережі до різнобічної обробки 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 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

KhNUE Simon Kuznets

S. KUZNETS KHNUE

Founder

Simon Kuznets Kharkiv National University of Economics, Nauka avenue, 9-A, Kharkiv, 61166, Ukraine
http://www.hneu.edu.ua/

Received on: 1st of December, 2018
Accepted on: 20th of December, 2018

© Oleg Rudenko, Oleksandr Bezsonov, Oleksandr Romanyk, 2018

Oleg Rudenko, Doctor of Science, Professor, Simon Kuznets Kharkiv National University of Economics, Ukraine.

Oleksandr Bezsonov, Doctor of Science, Professor, Simon Kuznets Kharkiv National University of Economics, Ukraine.

Oleksandr Romanyk, Ph.D. student, Kharkiv National University of Radio Electronics, Ukraine.

This is an Open Access article, distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

http://dx.doi.org/10.21511/dm.5(1).2019.03

відносяться штучні нейронні мережі. Результати імітаційного моделювання підтвердили, що запропонований метод навчання нейронної мережі дозволяє значно підвищити точність прогнозування часових рядів.

Ключові слова часовий ряд, штучна нейронна мережа, алгоритм навчання, активаційна функція, імітаційне моделювання

Класифікація JEL C45

ВСТУП

Прогнозування часових рядів (ЧР) 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 призвів 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. Серед існуючої безлічі статистичних моделей прогнозування найбільшою популярністю серед математиків, які займаються цим завданням, користується модель авторегресії і ковзного середнього (ARIMA). Своєю популярністю ця модель, в першу чергу, зобов'язана роботам Бокса і Дженкінса [7], які запропонували і розвинули узагальнену модель ARIMA. Але крім представлених цими авторами моделей, звичайно, існують і інші ефективні моделі і методи прогнозування, зокрема, авторегресійна модель проінтегрованого змінного середнього (ARIMAX). Слід також відзначити моделі експоненціального згладжування, які були запропоновані в роботах [8, 16] ще до появи робіт Бокса і Дженкінса.

Незважаючи на більш простий і доступний математичний апарат, прогнозування за допомогою моделей експоненціального згладжування часто дає результат, який можна порівняти з результатом, отриманим при використанні моделі ARIMA. Це не дивно, так як моделі експоненціального згладжування є окремим випадком моделей ARIMA. Іншими словами, кожна модель експоненціального згладжування має відповідну еквівалентну модель ARIMA.

До переваг авторегресійних моделей і методів, а також моделей і методів експоненціального згладжування можна віднести наступні:

1. Простота і прозорість моделювання.
2. Однаковість аналізу і проектування.

Слід зазначити, що цей клас моделей є одним з найбільш популярних, у зв'язку з чим існує багато прикладів їх застосування для вирішення задач прогнозування ЧР при вирішенні широкого спектру практичних задач. Відзначимо, що використання лінійних регресійних моделей забезпечує кращу швидкодію прогнозування, ніж використання інших моделей.

Недоліками даного класу моделей є:

1. Велике число параметрів моделі, ідентифікація яких неоднозначна.
2. Ресурсомісткість.
3. Лінійність та низька адаптивність моделей.
4. Відсутність здатності моделювання нелінійних процесів.

Основними недоліками нелінійних регресійних моделей є наступні:

http://dx.doi.org/10.21511/dm.5(1).2019.03

1. Складність визначення виду функціональної залежності.
2. Трудомісткість визначення параметрів моделі.

Відомо, що процес прогнозування в кожному конкретному випадку вимагає індивідуального підходу і зазвичай включає в себе цілий ряд процедур, наприклад:

1. Аналіз часовій послідовності на предмет наявності пропущених і випадають значень. Корекція цих значень.
2. Визначення наявності тренду і його типу. Визначення наявності періодичності в послідовності.
3. Перевірка послідовності на стаціонарність.
4. Аналіз послідовності на предмет необхідності попередньої обробки (логарифмування, взяття різниць тощо).
5. Вибір моделі.
6. Визначення параметрів моделі.
7. Дослідження адекватності моделі.
8. Прогноз на підставі обраної моделі.
9. Оцінка точності прогнозування моделі.
10. Аналіз характеру помилок обраної моделі.
11. Визначення адекватності обраної моделі і в разі необхідності її заміна і повернення до попередніх пунктів.

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.

1. ЛІТЕРАТУРНИЙ ОГЛЯД

Для вибору моделей, яка описують досліджуваний ЧР, були введені наступні кількісні метрики під назвою інформаційні критерії. Найбільш часто на практиці застосовуються інформаційний критерій Акаіке (AIC) і Байєсівський інформаційний критерій (або критерій Шварца (BIC або SIC)) [36].

Слід, однак, відзначити, що в більшості 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.

Альтернативою статистичним методам можуть служити методи обчислювального інтелекту, до числа яких, в першу чергу, слід віднести штучні нейронні мережі (ШНМ) [6, 21, 29, 34, 36]. 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. 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.

Слід зазначити, що ці властивості ШНМ обумовили їх досить широке використання при рішенні різноманітних задач прогнозування. Так у загальній проблемі нейромережевого прогнозування особливе місце займає задача прогнозування споживання електричної енергії [4, 9, 13, 19, 20, 24, 31, 35, 39], завдяки

своїй практичній значущості і складності (хаотичність, квазіперіодичність, скачки) відповідних ЧР, структура яких може змінюватися непередбачуваним чином.

В роботах [15, 28] розглядалося використання ШНМ для рішення задач прогнозування короткострокового навантаження.

В роботах [3] та [22] наводилися дослідження щодо використання ШНМ до прогнозування стоку ріки та телекомунікаційного трафіку відповідно.

В роботі [40] описано застосування ШНМ для прогнозування рівня глюкози в крові для хворих на інсулінозалежний діабет.

Існує досить багато робіт присвячених нейромережевому прогнозуванню в бізнесі, зокрема [1, 2, 11, 17, 18, 25, 30, 32, 37, 41].

Ще більше робіт присвячено висвітленню загальних питань використання штучних нейронних мереж в задачах прогнозування, наприклад [10, 27]. Слід зазначити, що в роботі [10] розглядається важливе питання робастного нейропрогнозування.

Аналіз особливостей нейропрогнозування дозволяє зробити такий висновок: у загальному випадку, завдання прогнозування часових рядів за допомогою ШНМ зводиться до наступної послідовності етапів:

1. Збір даних для навчання (етап, який вважається одним з найбільш складних).
2. Підготовка і нормалізація даних (приведення до виду ЧР).
3. Вибір топології ШНМ (на цьому етапі приймається рішення про кількість шарів і наявності зворотного зв'язку).
4. Емпіричний (шляхом експерименту) підбір характеристик ШНМ.
5. Емпіричний підбір параметрів навчання.
6. Навчання ШНМ.
7. Перевірка навчання на адекватність поставленим завданням.
8. Коригування параметрів з урахуванням попереднього кроку, остаточне навчання.
9. Вербалізація ШНМ (мінімізація опису з використанням декількох алгебраїчних або логічних функцій) з метою подальшого використання.

Таким чином, використання ШНМ вимагає вирішення задач структурної та параметричної оптимізації, відповідних вибору оптимальної топології мережі і її навчання.

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

2. МЕТА ДОСЛІДЖЕННЯ

Метою роботи є розробка нейромережевих методів та засобів для прогнозування часових рядів у реальному часі шляхом синтезу обчислювальних структур, що дозволять підвищити якість прогнозування при збереженні або покращанні функціональних характеристик, на основі штучних нейронних мереж.

3. МЕТОДИ ДОСЛІДЖЕННЯ

Методи дослідження базуються на теорії обчислювального інтелекту, а саме на методах теорії штучних нейронних мереж, які дозволили синтезувати нейромережеві моделі та отримати процедури їх навчання.

http://dx.doi.org/10.21511/dm.5(1).2019.03

4. БАГАТОШАРОВИЙ ПЕРСЕПТРОН ЯК НЕЙРОМЕРЕЖЕВА МОДЕЛЬ

Метою будь-якого прогнозування є створення моделі, яка дозволяє прогнозувати майбутнє і оцінити тенденції в змінах того чи іншого фактору. 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.

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

де $x = [x_1, x_2 ..., x_N]^T$ - вектор узагальненого вхідного сигналу Nx1; $f(\bullet)$ - 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 не лінійності 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 $\{x(k), y(k)\}, k = 1, 2, \ldots$.

Серед найбільш широко застосовуваних для вирішення цієї задачі ШНМ (багатошаровий персептрон (БП), радіально-базисні мережі (РБМ) і мережа CMAC (Cerebellar Model Articulation Controller) [6, 21, 29, 34, 36]), досить ефективним є БП, який використовує апроксимації нелінійного оператора $f(\bullet)$ виду

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

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:

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

та

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

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» [6, 21, 24, 28, 29], нейронна мережа отримує на вхід разом з поточним значенням затримані в часі минулі вхідні сигнали. 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 [28], 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 (мережі Джордана, Елмана тощо) [6, 21, 29, 34, 36]. Слід відзначити, що ці мережі потенційно придатні для прогнозування (зокрема, часових рядів). Однак відомо також, що нейронні мережі прямого поширення з успіхом справляються з цим завданням - правда, не у всіх випадках. При цьому застосовуються спеціальні методики «рекурентного навчання в реальному часі» або «зворотне поширення в часі» для визначення впливу минулих епох навчання на результат, отриманий на поточній ітерації. Налаштування ваг нейромережі здійснюється також будь-яким градієнтним алгоритмом. Структура таких мереж більш відповідає динамічним процесам, що моделюються, і тому їх застосування при вирішенні задач керування і багатокрокового прогнозування є більш ефективним. Але слід відзначити, що навчання таких мереж є складнішим завданням у зв'язку з наявністю в них додаткових ступенів свободи. Питання сталості таких мереж також необхідно досліджувати. Це пов'язано з тим, що при розрахунках динамічних похідних у персептроно-подібних мережах може мати місце ефект зникнення градієнта [21].

5. ВИБІР СТРУКТУРИ МЕРЕЖІ

Кількість нейронів прихованого шару попередньо можна оцінити таким чином [23]:

$$\frac{N}{10}-n-m \le n_k \le \frac{N}{2}-n-m, \qquad (5)$$

де n_k - кількість нейронів в прихованому шарі; n - кількість вхідних сигналів; m - кількість вихідних сигналів; N - кількість елементів статичної вибірки, необхідної для навчання.

З іншого боку, значення n_k можна оцінити таким чином [26]:

$$\frac{N\cdot\varepsilon_i}{n} > n_k, \qquad (6)$$

де ε_i - відносна помилка програмування нейромережі.

З урахуванням (6) можна записати:

$$\begin{cases} \frac{N}{10}-n-m \le \frac{N\cdot\varepsilon_i}{n}; \\ \frac{N\cdot\varepsilon_i}{n} \le \frac{N}{2}-n-m. \end{cases} \qquad (7)$$

Вирішуючи (7) відносно N, отримуємо:

$$\frac{2n(n+m)}{(n-10\varepsilon_i)} \le N \le \frac{10n(n+m)}{(n-10\varepsilon_i)}. \qquad (8)$$

При $\varepsilon_i \to 0$ нерівність (8) приймає вид:

$$2n(n+m) \le N \le 10n(n+m). \qquad (9)$$

Кількість синаптичних ваг для персептрона с одним прихованим шаром обчислюється так:

$$K_w = n\cdot n_k \cdot m. \qquad (10)$$

http://dx.doi.org/10.21511/dm.5(1).2019.03

6. АЛГОРИТМ НАВЧАННЯ МЕРЕЖІ

Розглянемо задачу прогнозування часового ряду, що описується рівнянням:

$$y(k) = f\left(z(k)\right) = f\left(w^{*T} x(k)\right), \qquad (11)$$

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:

$$e(k) = y^*(k) - y(k), \qquad (12)$$

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

Точність прогнозування будемо оцінювати за критерієм:

$$E(k) = 0,5e^2(k). \qquad (13)$$

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

$$w(k+1) = w(k) + \Delta w(k), \qquad (14)$$

де $\Delta w(k) = -\gamma \nabla_w [E(k)]_{w=w(k)}$ – 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 обмежимось лінійними членами:

$$e(k+1) = e(k) + \left(\frac{\partial e(k)}{\partial w(k)}\right)^T \Delta w(k) + \left(\frac{\partial e(k)}{\partial x(k)}\right)^T \Delta x(k) + \frac{\partial e(k)}{\partial y^*(k)} \Delta y^*(k) + \ldots. \qquad (16)$$

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

$$e^2(k+1) = e^2(k) + 2e(k)\left(\frac{\partial e(k)}{\partial w(k)}\right)^T \Delta w(k) + \left(\frac{\partial e(k)}{\partial w(k)}\right)^T \Delta w(k)^2 \qquad (17)$$

Враховуючи (11), (12) можна отримати наступні вирази для похідних у рівнянні (17):

$$\frac{\partial e(k)}{\partial w(k)} = -f'\left(z(k)\right) x^T(k), \qquad (18)$$

а

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

Це дозволяє записати (15) наступним чином

$$\Delta w(k) = -\gamma e(k) f'\left(z(k)\right) x(k). \qquad (19)$$

Врахувавши, що:

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

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

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

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

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

Для збіжності алгоритму (14),(15) необхідне виконання нерівності:

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

звідки маємо

$$0<\gamma<\frac{2}{(f'\left(z(k)\right))^2x(k)^2}. \qquad (20)$$

з умови:

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

Можна отримати вираз для оптимального значення параметра γ, при якому швидкість збіжності алгоритму навчання буде максимальною:

$$\gamma^{OPT}(k)=\frac{1}{(f'\left(z(k)\right))^2x(k)^2}, \qquad (21)$$

Тобто отримуємо алгоритм Качмажа (Уідроу-Хоффа) [5-8].

При застосуванні АФ виду (3), (4) похідні, що використовуються в (20), (21) легко обчислюються за формулами (для нашого випадку):

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

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

або

$$f'_{th}\left(w^Tx\right)=x(k)\left(1-f^2{}_{th}\left(w^Tx\right)\right),$$

$$f'_{\log}\left(w^Tx\right)=x(k)f_{\log}\left(z(k)\right)(1-f_{\log}\left(w^Tx\right).$$

Для підвищення обчислювальної стійкості процедури навчання слід скористатися регуляризацією алгоритму навчання [12, 33, 38], тобто замість (21) взяти:

$$\gamma'(k)=\frac{1}{(f'\left(z(k)\right))^2x(k)^2+\delta(k)}, \qquad (22)$$

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

http://dx.doi.org/10.21511/dm.5(1).2019.03

По аналогії з (20) неважко отримати, що для збіжності регуляризованого алгоритму необхідне виконання нерівності:

$$0 < \frac{1}{(f'(z(k)))^2 x(k)^2 + \delta(k)} < \frac{2}{(f'(z(k)))^2 x(k)^2}. \qquad (23)$$

З нерівності (23) отримуємо, що регуляризований алгоритм буде збігатися, якщо параметр параметр регуляризації $\delta(k)$ задовольняє нерівності:

$$\delta(k) > -\frac{1}{2}(f'(z(k)))^2 x(k)^2. \qquad (24)$$

Як видно з (24), параметр $\delta(k)$ є змінним у часі і може коригуватися на кожному кроці при надходженні нової інформації, наприклад, таким чином:

$$\delta(k) = \delta(k-1) - \alpha \nabla_{\delta}[E(k)]_{\delta=\delta(k)}. \qquad (25)$$

але при цьому виникає питання щодо вибору параметра α.

Деякі рекомендації щодо вибору δ можна знайти в [5, 12, 33, 38].

7. МОДЕЛЮВАННЯ

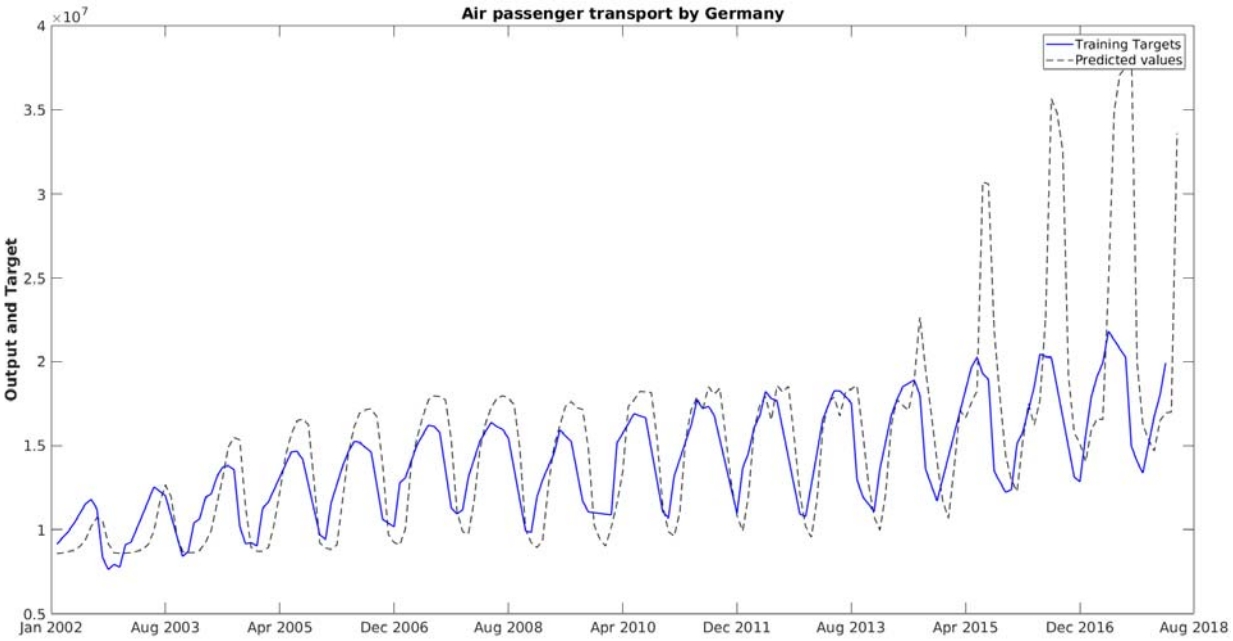
Розглядалась задача прогнозування об'єму пасажиропотоку у аеропортах Німеччини. Для цього було використано данні з офіційного сайту Eurostat [14] - статистичної служба Європейського союзу, що займається збором статистичної інформації по країнам-членам ЄС і гармонізацією статистичних методів використовуваних даними країнами.

Слід зазначити, що часові ряди, які відображають пасажиропотік будь-якого транспорту є надзвичайно популярні при тестуванні різноманітних систем прогнозування, оскільки вони мають усі необхідні складові часового ряду – тренд, сезонність та циклічність. Отже, для прогнозування об'єму пасажиропотоку у аеропортах Німеччини використовувався багатошаровий персептрон (БП) з трьома входами та одним виходом (для прогнозування кількості пасажирів у наступному місяці використовувались данні за три минулих місяця), БП мав два прихованих шара по 20 нейронів кожний. Для навчання нейронної мережі було сформовано навчаючу вибірку, у якій було відображено зміни пасажиропотоку Німеччини за останні 16 років. Результати моделювання з використанням різних значень параметру δ у алгоритмі навчання мережі наведено на Рисунку 1-а,б). На рисунках сполошною лінією показано реальний пасажиропотік, а пунктирною – спрогнозований за допомогою БП. Результати на Рисунку 1-а відповідають налаштуванню відповідно з (25) при $\alpha = 0.01$, а на Рисунку 1-б) – при $\alpha = 0.05$.

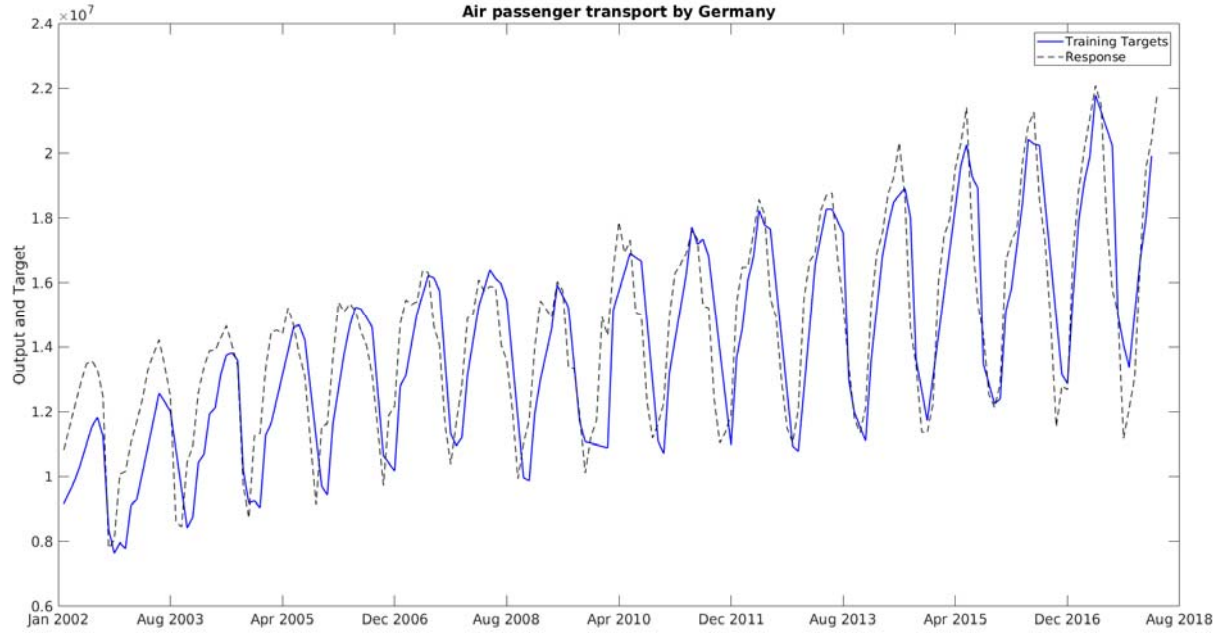
Як видно з результатів моделювання, при виборі значення параметру $\alpha = 0.01$ наприкінці часового ряду помилка прогнозування істотно зросла і результати прогнозування істотно відрізнялись від реальних значень. При виборі ж параметру $\alpha = 0.05$ вдалося отримати результати прогнозування, які суттєво не відрізнялись від реальних.

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.



а)



б)

Рисунок 1. Результати моделювання

Джерело: Розроблено авторами.

Наведено результати моделювання, які підтвердили ефективність використання регуляризованого алгоритму для навчання багатошарового персептрону зі змінним параметром регуляризації відповідно з формулою (25). Однак слід зазначити, що відкритим залишається питання щодо вибору оптимального значення іншого параметру (α), який входить в формулу для корекції параметра регуляризації.

http://dx.doi.org/10.21511/dm.5(1).2019.03

З метою підвищення якості прогнозування використовують рекурентні нейронні мережі. Структура цих мереж містить додаткові зворотні зв'язки, а для їх навчанні використовується специфічний алгоритм, який дозволяє розрахувати вплив власних похибок на якість прогнозу. Слід зазначити, що рекурентна нейронна мережа NARX (Nonlinear AutoRegression with eXternal inputs) є модифікацією багатошарового персептрона з додаванням зворотних зв'язків до вхідного шару нейронів, реалізованих затримками. Тому для її навчання можна використовувати досліджений в роботі алгоритм.

СПИСОК ЛІТЕРАТУРИ

1. Abbas, O. M. (2015). Neural networks in business forecasting. *International journal of computer*, *19*(1), 114-128. Retrieved from http://ijcjournal.org/index.php/InternationalJournalOfComputer/article/view/483
2. Abbas, O. M. (2017). *Business forecasting among neural networks and statistical methods* (120 p.). LAP LAMBERT Academic Publishing.
3. Amir, F. A., & Samir, I. S. (1999). A comparison between neural-network forecasting techniques – case study: river flow forecasting. *IEEE Transactions on neural networks, 10*(2), 402-409. https://doi.org/10.1109/72.750569
4. Amjady, N., & Keynia, F. (2011). A new neural network approach to short term load forecasting of electrical power systems. *Energies*, *4*(3), 488-503. https://doi.org/10.3390/en4030488
5. Benesty, J., & Paleologu, C. (2011). On regularization in adaptive filtering. *IEEE Transactions on audio, speech, and language processing, 19*(6), 1734-1742. Retrieved from http://externe.emt.inrs.ca/users/benesty/papers/aslp_aug2011.pdf
6. Bodyanskiy, Y. V., & Rudenko, O. G. (2004). *Искусственные нейронные сети: архитектура, обучение, применение [Iskusstvennyye neyronnyye seti: arkhitektura, obucheniye, primeneniye]* (372 p.). Kharkov: TELETEKH.
7. Boks, D., & Dzhenkins, H. (1974). *Анализ временных рядов прогноз и управление [Analiz vremennykh ryadov, prognoz i upravleniye]* (406 p.). Moskva: Mir.
8. Brown, R. G. (1959). *Statistical forecasting for inventory control* (232 p.). New York: McGraw-Hill.
9. Chen, J.-F., Do, Q. H., Nguyen T. V. A., & Doan T. T. H. (2018). Forecasting monthly electricity demands by wavelet neuro-fuzzy system optimized by heuristic algorithms. *Information, 9*(31), 15. https://doi.org/10.3390/info9030051
10. Chernodub, A. M. (2012). Навчання рекурентних нейронних мереж методом псевдорегуляризації для багатокрокового прогнозування часових рядів [Navchannia rekurentnykh neironnykh merezh metodom psevdorehuliaryzatsii dlia bahatokrokovoho prohnozuvannia chasovykh riadiv]. *Matematychni mashyny i systemy, 4*, 41-51. Retrieved from http://nbuv.gov.ua/UJRN/MMS_2012_4_6
11. Dalrymple, D. J. (1987). Sales forecasting practices: results of a united states survey. *International journal of forecasting, 3*, 379-392.
12. Demidenko, Y. Z. (1981). *Линейная и нелинейная регрессии [Lineynaya i nelineynaya regressii]* (302 p.). Moskva: Finansy i statistika.
13. Dogan, E., & Akgungor, A. (2013). Forecasting highway casualties under the effect of railway development policy in Turkey using artificial neural networks. *Neural computing and applications, 22*, 869-877. https://doi.org/10.1007/s00521-011-0778-0
14. Eurostat (n.d.). Retrieved from http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=avia_paoc&lang=en
15. Hippert, H. S., Pedreira, C. E., & Souza, R. C. (2001). Neural networks for short-term load forecasting: a review and evaluation. *IEEE Transactions on power systems, 16*(1), 44-55. https://doi.org/10.1109/59.910780
16. Holt, C. C. (1957). *Forecasting seasonal and trends by exponentially weighted moving averages* (ONR Memorandum No. 52) (236 p.). Pittsburgh: Carnegie institute of technology.
17. Hoptroff, R. G. (1993). The principles and practice of time series forecasting and business modeling using neural nets. *Neural computing & applications*, *1*(1), 59-66. https://doi.org/10.1007/BF01411375
18. Jerome, T. C., Douglas, R. M, & Atlas, L. E. (1994). Neural networks and robust time series prediction. *IEEE Transactions on neural networks, 5*(2), 240-254. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.108.6460&rep=rep1&type=pdf
19. Kandananond, K. (2011). Forecasting electricity demand in thailand with an artificial neural network approach. *Energies, 4*(8), 1246-1257. https://doi.org/10.3390/en4081246
20. Kazemi, S. M. R., Hoseini, S., Abbasian-Naghneh, M. M., & Habib, A. R. (2014). An evolutionary-based adaptive neuro-fuzzy inference system for intelligent short-term load forecasting. *International transactions in operational research, 21*(2), 311-326. https://doi.org/10.1111/itor.12046
21. Khaykin, S. (2006). *Нейронные сети: полный курс [Neyronnyye seti: polnyy kurs]* (1104 p.). Moskva: Vilyams.
22. Klevecka, I. (2008). Pre-processing of input data of neural networks: the case of forecasting telecommunication network traffic. *Telektronikk, 3*(4), 168-178.
23. Kruglov, V. V., & Borisov, V. V. (2002). *Искусственные нейронные сети. Теория и практика [Iskusstvennyye neyronnyye seti. Teoriya i praktika]* (382 p.). Moskva: Goryachaya liniya-Telekom.
24. Ling, S. H., Leung, F. H. F., Lam, H. K., & Tam, P. K. S. (2003). Short-term electric load forecasting based on a neural fuzzy network. *IEEE Transactions on industrial electronics*, *50*(6), 1305-1316. https://doi.org/10.1109/TIE.2003.819572
25. Maithili, A., Kumari R. V., & Rajamanickam S. (2012). Neural network towards business forecasting. *IOSR Journal of engineering, 2*(4), 831- 836. Retrieved from https://pdfs.semanticscholar.org/9b1f/23563d9c048ebbb4c93d4c6e0cdcfacd08e9.pdf
26. Makhotilo, K. V. (1998). *Разработка методик эволюционного синтеза нейросетевых компонентов систем управления [Razrabotka metodik evolyutsionnogo sinteza neyrosetevykh komponentov sistem upravleniya]* (179 p.). Kharkov: Kharkovskiy gosudarstvennyy politekhnicheskiy universitet.
27. Mandic, D. P., & Chambers, J. A. (2001). *Recurrent neural networks for prediction: learning algorithms, architectures and stability* (285 p.). England: John Wiley & Sons. Retrieved from https://doc.lagout.org/science/0_Computer%20Science/3_Theory/Neural%20Networks/Recurrent%20Neural%20Networks%20for%20Prediction.pdf

28. Mohsen, H., & Yazdan, S. (2007). Artificial neural network approach for short term load forecasting for Illam region. *World academy of science, engineering and technology*, *1*(4), 667-671. Retrieved from https://waset.org/publications/554/artificial-neural-network-approach-for-short-term-load-forecasting-for-illam-region-
29. Osovskiy, S. (2002). *Нейронные сети для обработки информации [Neyronnyye seti dlya obrabotki informatsii]* (344 p.). Moskva: Finansy i statistika.
30. Palit, A. K., & Popovic D. (2005). *Computational intelligence in time series forecasting: theory and engineering applications* (372 p.). Berlin: Springer.
31. Pliss, I. P., Popov, S. V., & Rybalchenko, T. V. (2008). Нейросетевое прогнозирование нестационарных временных рядов энергопотребления в условиях структурной неопределенности [Neyrosetevoye prognozirovaniye nestatsionarnykh vremennykh ryadov energopotrebleniya v usloviyakh strukturnoy neopredelennosti]. *Svitlotekhnika ta elektroenerhetyka, 3*, 41-48.
32. Raghupathi, W., Schkade L., & Bapi, R. (1991). A neural network application for bankruptcy prediction. In *Proceedings of the 24th hawaii international conference on system sciences*, *4*, 147-155.
33. Raybman, N. S., & Chadeyev, V. M. (1966). *Адаптивные модели в системах управления [Adaptivnyye modeli v sistemakh upravleniya]* (159 p.). Moskva: Sovetstkoye radio.
34. Rudenko, O. H., & Bodyanskyy, E. V. (2002). *Основы теории искусственных нейронных сетей [Osnovy teorii iskusstvennykh neyronnykh setey]* (317 p.). Kharkov: TELETEKH.
35. Shumilova, G. P., Gotman, N. E., Startseva, T. B. (2008). *Прогнозирование электрических нагрузок при оперативном управлении электроэнергетическими системами на основе нейросетевых структур [Prognozirovaniye elektricheskikh nagruzok pri operativnom upravlenii elektroenergeticheskimi sistemami na osnove neyrosetevykh struktur]* (77 p.). Syktyvkar: KNTS UrO RAN. Retrieved from http://energy.komisc.ru/downloads/docs/prognozirovanie_jelektricheskih_nagruzok_pri_operativnom_upravlenii_jelektrojenergeticheskimii_sistemami_na_osnove_nejrosetevyh_struktur.pdf
36. Snytyuk, V. E. (2008). *Прогнозирование. Модели, методы, алгоритмы [Prognozirovaniye. Modeli, metody, algoritmy]* (364 p.). Kiyev: Maklaut.
37. Tam, K. Y. (1991). Neural network models and the prediction of bank bankruptcy. *Omega, The international journal of management science*, 19, 429-445. https://doi.org/10.1016/0305-0483(91)90060-7
38. Tikhonov, A. N., & Arsenin, V. Y. (1980). *Методы решения некорректных задач [Metody resheniya nekorrektnykh zadach]* (223 p.). Moskva: Nauka.
39. Tzafestas, S., & Tzafestas, E. (2001). Computational intelligence techniques for short-term electric load forecasting. *Journal of intelligent and robotic systems*, *3*, 7-68. https://doi.org/10.1023/A:1012402930055
40. Yemelyanov, A. O., & Chernetsov, S. A. (2010). Нейросетевое прогнозирование уровня глюкозы в крови для больных инсулинозависимым диабетом [Neyrosetevoye prognozirovaniye urovnya glyukozy v krovi dlya bolnykh insulinozavisimym diabetom]. In *XII Vserossiyskaya nauchno-tekhnicheskaya konferentsiya «Neyroinformatika-2010»* (pp. 326-334). Moskva: NIAU MIFI.
41. Zhang, G. P. (Ed.) (2004). *Neural networks in business forecasting* (350 p.). Hershey: Idea Group Publishing.

http://dx.doi.org/10.21511/dm.5(1).2019.03