

О.М. Кісельова, О.М. Притоманова, Я.Б. Дон
Дніпровський національний університет імені Олеся Гончара

РОЗРОБКА НЕЙРОНЕЧІТКОЇ МОДЕЛІ КОРОТКОСТРОКОВОГО ПРОГНОЗУ КІЛЬКОСТІ ВИПАДКІВ ЗАХВОРЮВАНOSTI НА COVID-19

Розроблено нейронечітку модель прогнозування кількості випадків захворюваності на Covid-19. Перевагою побудованої моделі є можливість короткострокового прогнозу в умовах нечіткої як кількісної, так і якісної інформації за відносно малими експериментальними вибірками.

Ключові слова: часові ряди, короткостроковий прогноз, нейронечіткі технології, r -алгоритм Шора, негладкі оптимізаційні задачі.

E.M. Kiseleva, O.M. Prytomanova, Y.B. Don
Oles Honchar Dnipro National University

DEVELOPMENT OF NEURO-FUZZY MODEL FOR SHORT-TERM FORECASTING CONFIRMED CASES OF COVID-19

The article presents the development of a fuzzy model for forecasting the number of cases of Covid-19.

Known models for forecasting time series are built, as a rule, on the basis of the classical econometrics apparatus. However, to use them, it is necessary to have a significant amount of clear input statistical information, which is not always available and/or has a high degree of reliability. In addition, econometric models assume that the type of dependence sought can be linear, logarithmic, polynomial, and others.

The non-triviality of modeling such a complex process as the number of cases of COVID-19 is due to a number of factors, such as the lack of statistical information in the first days of the spread of the disease, its inaccuracy, and sometimes even contradiction. All this does not allow to ensure a high level of adequacy of the model which is built on the basis of the econometric approach.

Because of the above, in this work the modeling of the forecast of the number of cases of Covid-19 is based on neuro fuzzy technologies, which, on the one hand, are devoid of these shortcomings of econometric approaches, on the other hand, allows to expand possibilities of modeling of difficult objects, processes. This is very relevant in the real conditions of morbidity in the absence of reliable data, incomplete and fuzzy statistical information about the object, complex nonlinear dependencies of outputs on system inputs.

Neural fuzzy technologies are based on a combination of two fundamentally different mathematical constructs: neural networks and fuzzy logic. In recent years, neural networks have been used in practice wherever it is necessary to solve problems of forecasting, classification, and management. However, training neural networks requires a large sample of experimental data. Unlike neural networks, fuzzy technologies (fuzzy-systems) use expert information about the patterns that are found in the available experimental data, in the form of "if-then" language rules. These rules, which are formalized using fuzzy logic,

make it possible to build forecast models for relatively small (in comparison with neural networks) samples of experimental data. The combination of the two described technologies is a combination of neural network learning abilities with clarity and easy interpretation of fuzzy “if-then” rules. The Shor's r -algorithm method of non-differential optimization was used to adjust the fuzzy model.

The advantage of the constructed model is the possibility of short-term forecasting in conditions of fuzzy quantitative and qualitative information on relatively small experimental samples.

Keywords: Time series, short-term forecast, neuro-fuzzy technologies, Shor's r -algorithm, nonsmooth optimization problems.

Е.М. Киселева, О.М. Притоманова, Я.Б. Дон
Дніпровський національний університет імені Олеся Гончара

РАЗРАБОТКА НЕЙРОНЕЧЕТКОЙ МОДЕЛИ КРАТКОСРОЧНОГО ПРОГНОЗА КОЛИЧЕСТВА СЛУЧАЕВ ЗАБОЛЕВАНИЯ COVID-19

Построена нейронечеткая модель прогнозирования количества случаев заболеваемости Covid-19. Преимуществом построенной модели является возможность краткосрочного прогноза в условиях нечеткой как количественной, так и качественной информации по относительно малым экспериментальным выборкам.

Ключевые слова: временные ряды, краткосрочный прогноз, нейронечеткие технологии, r -алгоритм Шора, негладкие оптимизационные задачи.

Вступ. Прогнози, які побудовані на основі застосування математичних методів та моделей, поділяють на казуальні й неказуальні [1]. Найпоширенішими серед неказуальних моделей є моделі часових рядів [2], в основі яких лежить перенесення минулих і теперішніх закономірностей, зв'язків і співвідношень на майбутнє. Приклади такого прогнозування є у багатьох вітчизняних та зарубіжних публікаціях. Але ці методи добре працюють, якщо система, яка породжує даний часовий ряд, лінійна або піддається формалізованому опису [6].

Останнім часом запропоновано нові підходи до побудови нелінійних моделей, які пов'язані з застосуванням теорії нечітких множин [3, 4] і нейронних мереж [5, 6]. Синтез двох принципово різних математичних конструкцій нейронних мереж і нечіткої логіки є основою нейронечітких технологій [7, 8, 9]. Ефективність нейронечіткого підходу особливо виявляється тоді, коли процеси є досить складними для аналізу за допомогою класичних кількісних методів, або коли джерела інформації інтерпретуються якісно, неточно або невизначено [10-18].

У роботі застосовано нейронечіткі технології до розробки короткострокового прогнозу кількості випадків захворюваності на Covid-19.

Постановка задачі. Побудуємо нейронечітку модель прогнозу, яка ґрунтується на поданні вхідної статистичної інформації щодо динаміки кількості зареєстрованих випадків захворюваності на Covid-19 у вигляді нечіткої продукційної бази знань. У базі знань також буде подано експертно-лінгвістичну інформацію стосовно закономірностей в існуючих даних.

У табл. 1 наведено щоденні статистичні дані щодо кількості зареєстрованих випадків (значення змінної PR) за період з 17.04.2020 по 14.05.2020, де через x_1, x_2, \dots, x_{28} позначено кількість випадків за i -й день (період). Дані взято з офіційного сайту МОЗ.

На базі тільки цих даних нам необхідно побудувати модель прогнозування.

Матеріали і методи. Алгоритм побудови нейронечіткої моделі прогнозування складається з двох етапів [14]:

1) формування нечіткої бази знань про об'єкт і побудова на її основі нечіткої моделі об'єкта з кількома входами і одним виходом, яка грубо відтворює залежність виходу від входів за допомогою лінгвістичних правил «ЯКЦО-ТО», що генеруються з експериментальних даних;

2) настройка (оптимізація) параметрів нечіткої моделі: пошук таких параметрів нечіткої моделі, які мінімізують відхилення модельних значень від експериментальних.

Таблиця 1

Щоденні дані про кількість зареєстрованих випадків захворюваності на Covid-19 у Дніпропетровській області

Період	Кількість зареєстрованих випадків (PR)	Позначення	Період	Кількість зареєстрованих випадків (PR)	Позначення
17.04.2020	15	x_1	01.05.2020	72	x_{15}
18.04.2020	6	x_2	02.05.2020	12	x_{16}
19.04.2020	8	x_3	03.05.2020	12	x_{17}
20.04.2020	4	x_4	04.05.2020	18	x_{18}
21.04.2020	28	x_5	05.05.2020	90	x_{19}
22.04.2020	36	x_6	06.05.2020	56	x_{20}
23.04.2020	9	x_7	07.05.2020	28	x_{21}
24.04.2020	45	x_8	08.05.2020	29	x_{22}
25.04.2020	52	x_9	09.05.2020	23	x_{23}
26.04.2020	18	x_{10}	10.05.2020	27	x_{24}
27.04.2020	8	x_{11}	11.05.2020	22	x_{25}
28.04.2020	5	x_{12}	12.05.2020	23	x_{26}
29.04.2020	24	x_{13}	13.05.2020	5	x_{27}
30.04.2020	17	x_{14}	14.05.2020	10	x_{28}

Для реалізації першого етапу алгоритму, аналізуючи динаміку змін кількості зареєстрованих випадків захворюваності на Covid-19 у Дніпропетровській області (рис. 1), побудуємо нечітку базу знань. Для цього згрупуємо елемен-

ти часового ряду значень для змінної PR в цикли по сім елементів: три елементи на вхід, один - на вихід. Схематично це можна подати у вигляді:

$$x_1 x_2 \underbrace{x_3 x_4 x_5 x_6 x_7 x_8 x_9}_{3\text{-й цикл}} x_{10} \dots x_{28}, \quad (1)$$

Тобто ми будемо вважати, що прогноз значення змінної PR на i -й ($i \geq 8$) період (день) залежить від тенденцій, що склалися в попередніх – $(i-1)$ -у, $(i-2)$ -у та $(i-3)$ -у періодах. Отже, прогнозне значення змінної PR на i -й період є функція (поки ще невідома) параметрів $x_{i-3}, x_{i-2}, x_{i-1}$:

$$x_i = f(x_{i-3}, x_{i-2}, x_{i-1}), \quad i \geq 8. \quad (2)$$

З урахуванням цих припущень, закономірності, які можна простежити на графіку статистичних даних або сформулювати за допомогою думки експертів, запишемо у вигляді висловлювань на природній мові. Ці висловлювання є правилами типу ЯКЩО-ТО. Вони формують нечітку базу знань, наприклад:
 $i=8$: ЯКЩО $\langle x_{i-3} = \text{низький} \rangle \wedge \langle x_{i-2} = \text{середній} \rangle \wedge \langle x_{i-1} = \text{середній} \rangle$
 ТО $\langle x_i = \text{середній} \rangle$.

З бази знань зрозуміло, що при такому підході ми маємо можливість прогнозувати на 6-7 періодів (днів) вперед.

Для запису моделі (2) у явному вигляді лінгвістичні оцінки «низький», «високий» та інші формалізуємо за допомогою функцій належності.

Задамо ці функції у такому вигляді:

$$\mu^T(x) = \frac{1}{1 + \left(\frac{x-b}{c}\right)^2}, \quad (3)$$

де b та c – параметри, які спочатку обираються експертом, а потім настраюються за експериментальними даними: b – координата максимуму функції $\mu^T(x)$, причому $\mu^T(b) = 1$, c – коефіцієнт концентрації – розтягнення функції $\mu^T(x)$, $\mu^T(x)$ – число у діапазоні $[0,1]$, яке характеризує суб'єктивну міру відповідності значення x лінгвістичній оцінці T .

Далі, застосовуючи алгоритм нечіткого логічного виведення [14], будемо нечітку модель прогнозування, за допомогою якої можна грубо прогнозувати кількість випадків на 6-7-м кроків вперед. Треба відзначити, що кроків (періодів) прогнозування може бути і значно більше. Але у цьому разі ступінь достовірності прогнозу падає із збільшенням кількості кроків.

З метою покращення точності прогнозу необхідно настроїти (навчити) модель за експериментальними даними. Налаштування (оптимізація параметрів) моделі – це підбір таких параметрів b і c функцій належності (3), які забезпечать мінімум розходжень між модельними (теоретичними, одержаними з моделі) і експериментальними даними. Для налаштування нечіткої моделі було використано модифікацію градієнтного методу на випадок недиференційовних функцій, r -алгоритм Н.З. Шора [19, 20].

Перевірка побудованої моделі на адекватність проводилась різними способами. Один з них, наприклад, полягає в такому: налаштування моделі спочатку проводилась за експериментальними даними з 17.05.2020 по

10.05.2020. Тобто кількість експериментальних значень змінної PR дорівнювала 24. Потім прогнозні значення, одержані за моделлю, порівнювалось з фактичним. Отримані достатньо близькі (з точністю від 0.01 до 0.04) значення теорії та експерименту свідчать про високу адекватність побудованої моделі.

Результати. З метою реалізації розробленого алгоритму побудови нейронечіткої моделі прогнозу був створений пакет програм у середовищі VisualStudio 2019 мовою C++, у якому автоматизовано найбільш трудомістка операція по генерації правил для бази знань, також є можливість вносити зміни у згенеровані правила.

За допомогою пакету програмної реалізації алгоритму побудови та настройки нейронечіткої моделі проведено обчислювальні експерименти, які виявили, що модель дозволяє одержувати якісний прогноз у разі невеликих (порівняно із статистичними моделями) експериментальних вибірок. При збільшенні кількості вхідних точок збільшується кількість правил у базі знань. В той же час, при збільшенні кількості вхідних точок поліпшується якість прогнозу, а при збільшенні кількості правил у базі знань поліпшується якість настройки.

За допомогою пакету програмної реалізації розробленого алгоритму побудови нейронечіткої моделі прогнозу розраховані та представлені нижче короткострокові прогнози кількості випадків захворювання на Covid-19 у Дніпропетровській області, Дніпрі та Україні на період з 15.05.2020 по 21.05.2020.



Рис. 1. Кількість зареєстрованих випадків за день у Дніпропетровській області та прогноз захворюваності на наступні 7 днів

Прогноз будувався при припущенні, що періодичність даних складає 7 днів, причому на 8-й день (тобто на 1-й день наступного періоду) впливають дані за 1-й, 2-й та 3-й дні попереднього періоду.

Для періоду з 15.05.2020 по 21.05.2020 середній прогноз випадків захворювань за день у Дніпропетровській обл. складає 26 випадків (песимістичний прогноз – 69 випадків, оптимістичний прогноз – 11 випадків). Точність прогнозу 1,1%.



Рис. 2. Кількість зареєстрованих випадків за день у м. Дніпро та прогноз захворюваності на наступні 7 днів

Для періоду з 15.05.2020 по 21.05.2020 середній прогноз випадків захворювань за день у м. Дніпро складає 16 випадків (песимістичний прогноз – 33 випадки, оптимістичний прогноз – 5 випадків). Точність прогнозу 3,8%.

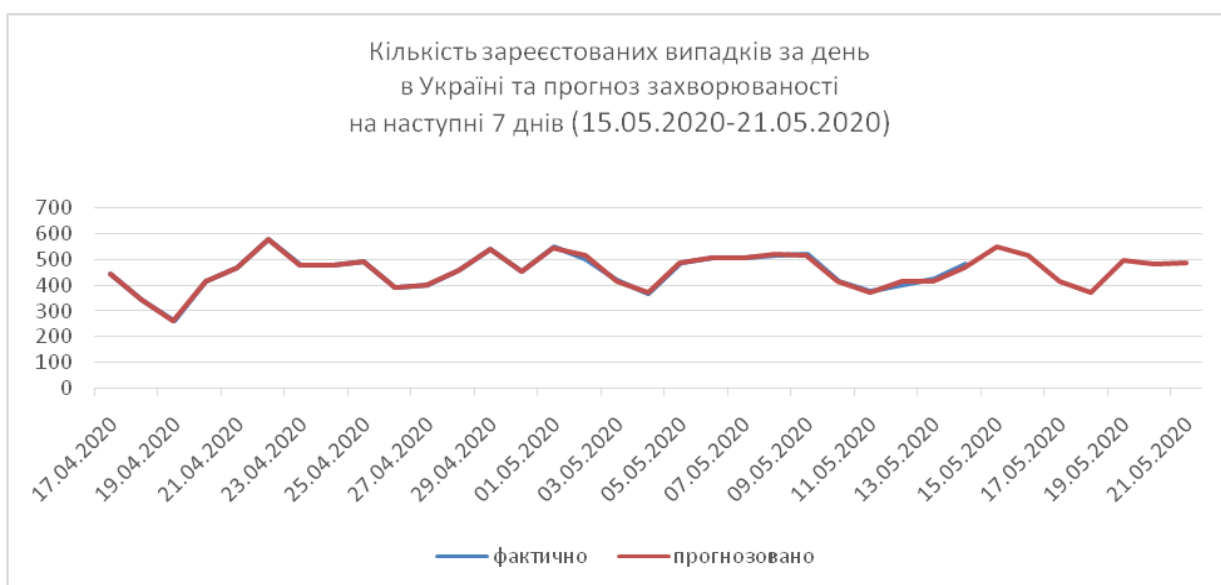


Рис. 3. Кількість зареєстрованих випадків за день в Україні та прогноз захворюваності на наступні 7 днів

Для періоду з 15.05.2020 по 21.05.2020 середній прогноз випадків захворювань за день в Україні складає 474 випадків (песимістичний прогноз – 550 випадків, оптимістичний прогноз – 370 випадків). Точність прогнозу 1,0%.

Прогноз на рис.1-3 був розроблений 14.05.2020 на наступні 7 днів (з 15.05.2020 по 21.05.2020 включно). Перевірка прогнозу була проведена 25.05.2020, коли стали відомі дані за період з 15.05.2020 по 21.05.2020.

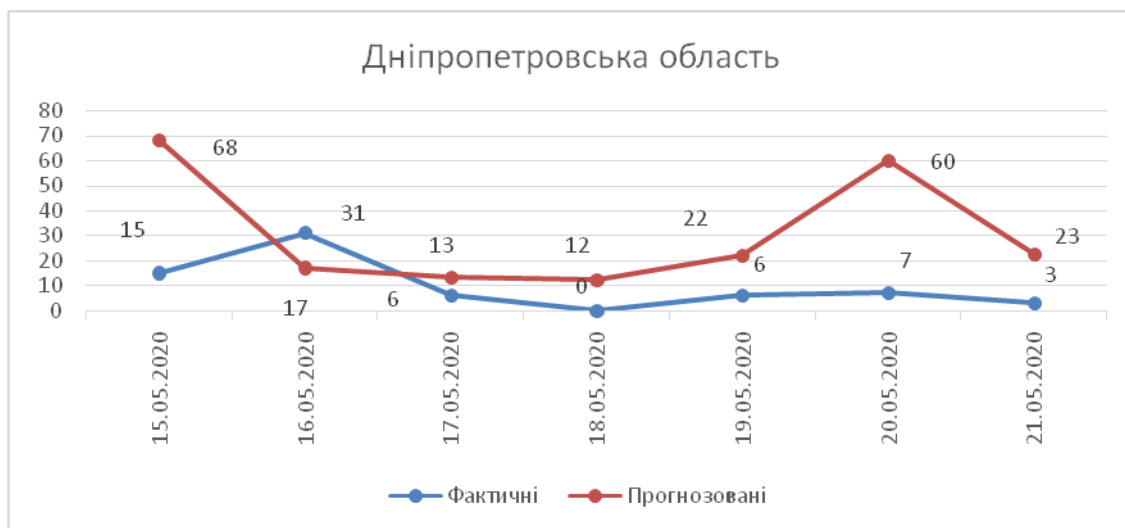


Рис. 4. Перевірка прогнозу. Дніпропетровська область. Середнє значення прогнозу 30 випадків. Середнє фактичне значення 10 випадків.



Рис. 5. Перевірка прогнозу, м. Дніпро. Середнє значення прогнозу 15 випадків. Середнє фактичне значення 10 випадків.

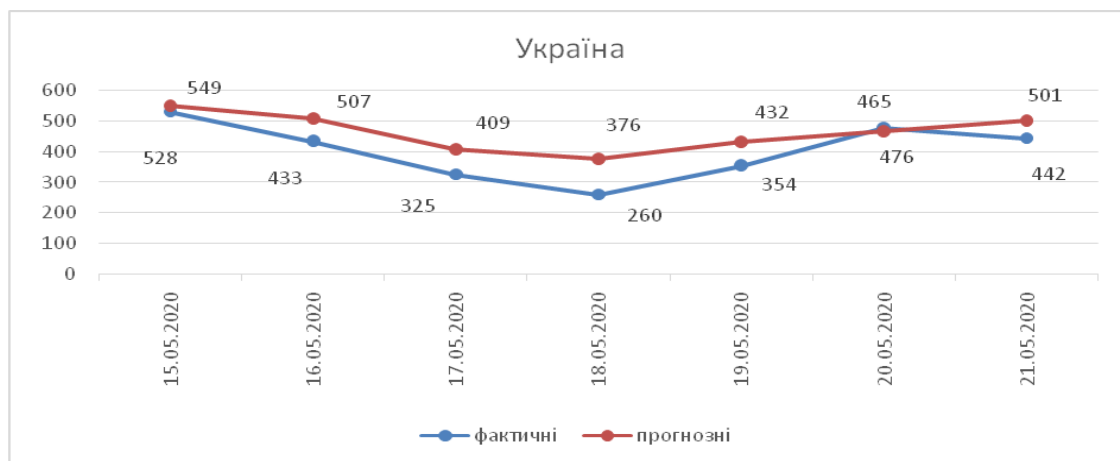


Рис. 6. Перевірка прогнозу. Україна. Середнє значення прогнозу 462 випадки. Середнє фактичне значення 403 випадки.



Рис. 7. Кількість зареєстрованих випадків за день у Дніпропетровській області та прогноз захворюваності на наступні 7 днів з 25.05.2020 по 31.05.2020: середній прогноз за день складає 24 випадки, песимістичний прогноз – 50 випадків, оптимістичний прогноз – 6 випадків. Точність прогнозу 1,7%.



Рис.8. Кількість зареєстрованих випадків за день у м. Дніпро та прогноз захворюваності на наступні 7 днів з 25.05.2020 по 31.05.2020: середній прогноз за день складає 11 випадків, песимістичний прогноз – 25 випадків, оптимістичний прогноз – 6 випадків. Точність прогнозу 1,8%.



Рис. 9. Кількість зареєстрованих випадків за день в Україні та прогноз захворюваності на наступні 7 днів з 25.05.2020 по 31.05.2020: середній прогноз за день складає 432 випадки, песимістичний прогноз – 528 випадків, оптимістичний прогноз – 372 випадки. Точність прогнозу 1,3%.

Прогноз на рис.7-9 був розроблений на наступні 7 днів (з 25.05.2020 по 31.05.2020 включно). Перевірка прогнозу була проведена 01.06.2020, коли стали відомі дані за період з 25.05.2020 по 31.05.2020.



Рис. 10. Перевірка прогнозу. Дніпропетровська область: Середнє значення прогнозу 24 випадки. Середнє фактичне значення 10 випадків.

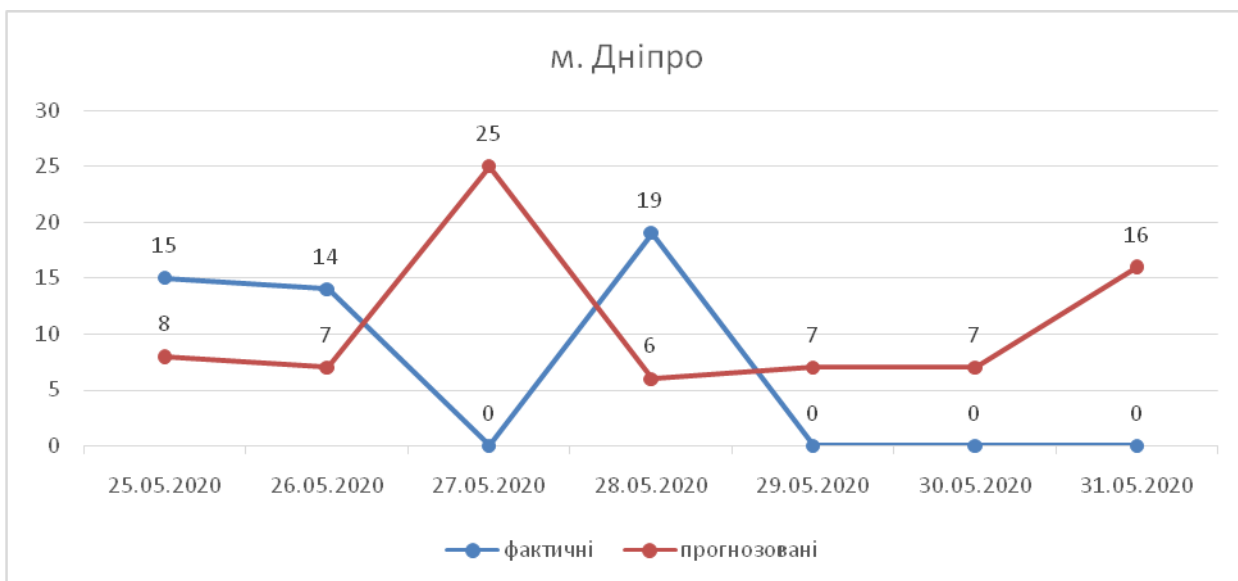


Рис. 11. Перевірка прогнозу. м. Дніпро:
Середнє значення прогнозу 11 випадків.
Середнє фактичне значення 7 випадків.

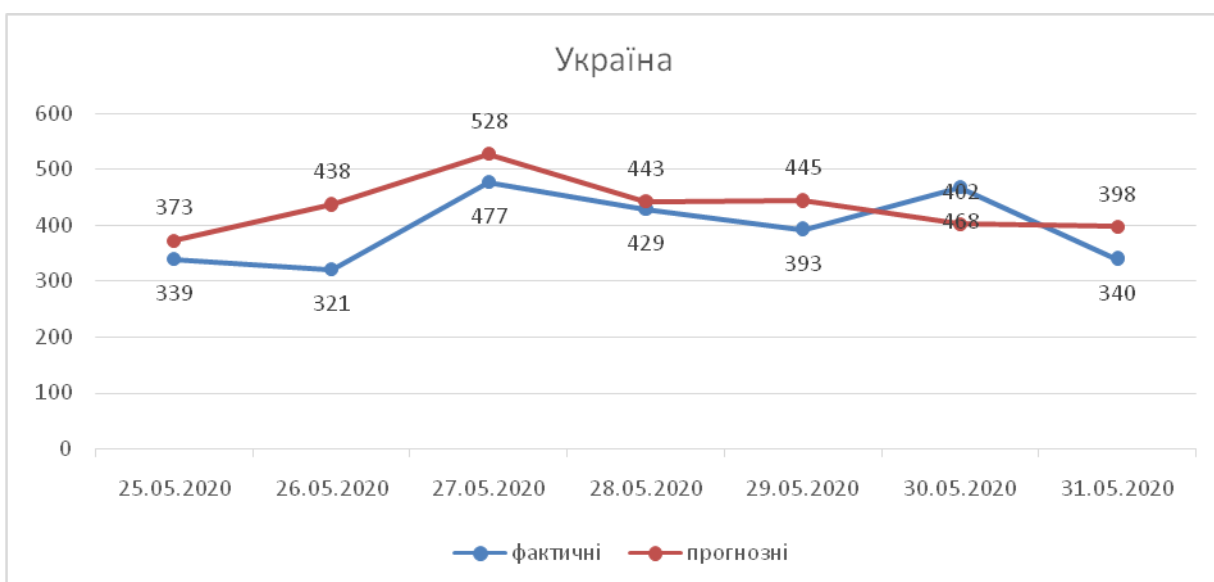


Рис. 12. Перевірка прогнозу. Україна:
Середнє значення прогнозу 432 випадки.
Середнє фактичне значення 395 випадків.

Висновки. Застосування нейронечіткого підходу до прогнозування таких досить унікальних процесів при невеликих вибірках статистичних даних (які можуть бути неточними) дозволяє виявляти тренди, а не надавати детальні щоденні дані з твердою гарантією їх виконання.

Бібліографічні посилання

1. **Gardner, E.S., Makridakis, S.** The Future of Forecasting [Text] // International Journal of Forecasting – 1988. – Vol. 4. – P. 325-330.

2. **Box, G.E., Jenkins, G.M., Reinsel, G.C.** Time Series Analysis: Forecasting And Control [Text] // WILEY.2008.
3. **Заде, Л.А.** Понятие лингвистической переменной и её применение к принятию приближенных решений [Текст]. – М.: Мир, 1976. – 167с.
4. **Zimmerman, H.-J.** Fuzzy Sets Theory – and Its Applications [Text] – 4th ed. - Boston: Kluwer Academic Publishers, 2001. - 514 p.
5. **Хайкин, С.** Нейронные сети: полный курс [Текст]. – М.: Вильямс, 2006. – 1104 с.
6. **Kocadagli, O., Askgil, B.** Nonlinear time series forecasting with Bayesian neural networks Expert [Text] // Systems with Applications. 2014 (41). – P. 6596–6610.
7. **Siddique, N.** Intelligent Control: A Hybrid Approach Based on Fuzzy Logic, Neural Networks and Genetic Algorithms [Text]/ – Springer, 2014. — 282 p.
8. **Касprzyk, J., Pedrycz, W.** Springer Handbook of Computational Intelligence. – Springer, 2015. — 1633 p.
9. **Борисов, В.В.** Нечеткие модели и сети [Текст] / В. В. Борисов, В. В. Круглов, А. С. Федулов. – М.: Горячая линия-Телеком, 2007. – 284 с.
10. **Nie, J.** Nonlinear time series forecasting: A fuzzy neural approach [Text] Neurocomputing. – 1999. – 16. N1. – P. 63-76.
11. **Ротштейн, О.П.** Интеллектуальные технологии идентификации: нечеткие множества, генетические алгоритмы, нейронные сети [Текст]. – УНІВЕРСУМ. – Вінниця, 1999. - 320 с.
12. **Кісельова, О.М., Бойцун, Н.Є., Притоманова, О.М.** Прогнозування економічних і фінансових процесів на основі нейронечітких технологій [Текст]. // Фінанси України, 2005. – №5. - С.87-99.
13. **Притоманова, О.М.** Моделювання кредитного ризику комерційного банку на основі нейронечітких технологій: автореф. дис...канд. екон. наук: спец. 08.00.11; Дніпр. нац. ун-т. – Д., 2004. – 19 с.
14. **Elena M. Kiseleva, Olga M. Prytomanova, Sergey V. Zhuravel** Valuation of Startups Investment Attractiveness Based on Neuro-Fuzzy Technologies [Text] // Journal of Automation and Information Sciences – 2016. – P. 1-22.
15. **Quek, C., Pasquier, M., Boon, B., Lim, S.** POP-TRAFFIC: A Novel Fuzzy Neural Approach to Road Traffic Analysis and Prediction [Text] // IEEE Transactions On Intelligent Transportation Systems, Vol. 7, No. 2, June 2006.
16. **Monadjemi,S.A., Moallem, P.** Automatic Diagnosis of Particular Diseases Using a Fuzzy-Neural Approach [Text] // International Review on Computers and Software. – Vol. 3. N. 4 July 2008.
17. **Kazarian, A., Kiseleva, E., Prytomanova, O., Timms, J.** Testing breast cancer serum biomarkers for early detection and prognosis in pre-diagnosis samples [Text] // British Journal of Cancer. 2017. №116. – P. 501-508.
18. **Blyuss, O., Kiseleva, E., Prytomanova O., Crnogorac-Jurcevic, T.** Development of PancRISK, a urine biomarker-based risk score for stratified screening of pancreatic cancer patients [Text] // British Journal of Cancer. 2020. Vol. 122(6). – P. 692–696.
19. **Кісельова, О.М., Притоманова, О.М., Падалко, В.Г.** Про оптимізацію параметрів нейронечіткої моделі експортних відносин між Україною та Китаєм [Текст] // Системний аналіз та інформаційні технології: матеріали 20-ї науково-технічної конференції SAIT 2018. – Київ:ННК «ІПСА» НТУУ «КПІ». 2018. – С. 68.
20. **Шор, Н.З.** Методы минимизации недифференцируемых функций и их приложения [Текст]. – Киев: Наук. Думка, 1979. – 199 с.

Надійшла до редколегії 23.09. 2020.