

АНОТАЦІЯ

Токарева К.А. Гібридні та ансамблеві методи та моделі машинного навчання прогнозування фінансових часових рядів. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 113 – Прикладна математика. – Чернівецький національний університет імені Юрія Федьковича МОН України, Чернівці, 2024.

Дисертаційна робота присвячена побудові гібридних алгоритмів на основі методів та моделей машинного навчання та класичних алгоритмів часових рядів.

Результати дисертаційної роботи є підґрунтям для подальших теоретичних і практичних наукових розробок у дослідженні гібридних алгоритмів побудови прогнозу для фінансових часових рядів.

Дисертація складається із вступу, чотирьох розділів, висновків, переліку використаних джерел та додатку.

У вступі обґрунтовано актуальність теми дослідження, сформульовано мету, завдання, предмет, об'єкт та методи дослідження, вказано наукову новизну, теоретичне та практичне значення отриманих результатів, проаналізовано зв'язок роботи з науковими дослідженнями та особистий внесок здобувача, а також наведено відомості про апробацію та публікації основних результатів дисертації. Описано структуру та обсяг дисертаційної роботи.

У першому розділі здійснено огляд наукової літератури, присвяченої основним напрямкам досліджень часових рядів в цілому та фінансових часових рядів зокрема, розглянуто ключові відомості з теорії часових рядів та методів машинного навчання, наведено опис основних напрямів досліджень та визначено завдання, якими займається теорія часових рядів включаючи основні проблеми побудови прогнозу фінансових часових рядів. Проведено детальний огляд основних класичних алгоритмів побудови прогнозу для часового ряду, зокрема, а також їх сфери застосування. Детально

проаналізовано хронологію розвитку теорії часових рядів та умовні періоди ускладнення моделей. У другому розділі даного розділу розглянуто основні моделі штучного інтелекту та машинного навчання для аналізу часових рядів, описано топології нейронних мереж, які будуть використовуватися при побудові гібридних алгоритмів.

У другому розділі проведено дослідження гібридних моделей, в яких основний процес оцінювався на основі моделі часового ряду, а залишки моделі оцінювалися на основі нейронних мереж. Використання такого підходу дало змогу розширення використання класичних моделей часових рядів на випадок відсутності гомоскедастичності у залишках та відсутності незалежності залишків часового ряду. А саме показано, що

- часові ряди фондових індексів містять як лінійну, так і не лінійну складову, а отже, окремо лінійна модель ARIMA та нелінійна модель ANN не можуть дати точну оцінку таких часових рядів;
- розроблено гібридну модель моделі авторегресії-ковзної середньої та штучної нейронної мережі прогнозування часових рядів фондових індексів;
- доведено, якщо часовий ряд є строго стаціонарним, то ARIMA модель є більш адекватною для гауссівських рядів в розробленій гібридній моделі.
- здійснено порівняльний аналіз запропонованого алгоритму із класичними алгоритмами оцінки прогнозу фінансових часових рядів, в результаті чого розроблений метод продемонстрував більшу точність порівняно з іншими гібридними моделями, оскільки ряд даних розкладається на компоненти з низькою та високою волатильністю (лінійну та нелінійну), а отже сумарна похибка моделі є низькою.

У третьому розділі розглянуто зв'язок хаотичних динамічних систем та часових. Проаналізовано основні підходи до визначення детермінованого хаосу та основні методи врахування наявності детермінованого хаосу при

оцінці прогнозу фінансового часового ряду. Для визначення існування хаосу в системі були представлені чотири методи:

- перетворення Фур'є;
- показник Ляпунова;
- показник Герста;
- фрактальна розмірність.

Основним підходом для застосування теорії хаосу, а саме детермінованого хаосу, є використання теореми вкладення Такенса, згідно якої структуру хаотичного детермінованого процесу можна спростити використовуючи розбиття векторного простору $(x(t), \dots, x(t + T))$. Метод реконструкції фазового простору використовується для реконструкції точок фазового простору з хаотичних часових рядів з використанням розмірності вкладення D та часової затримки T . Відповідно до методу реконструкції фазового простору, точки фазового простору відтворюють ключові характеристики вихідного часового ряду. Для підтвердження ефективності запропонованої методики в даному розділі було досліджено три різні класичні хаотичні часових ряди: Маккея-Глесса, логістичного та Ено. Запропонований метод об'єднувався з генетичними алгоритмами для покращення результатів прогнозування. Було показано, що новий метод прогнозування, покращений за допомогою GA, демонструє визначні результати. Порівняння результатів запропонованої методики з результатами, наведеними в літературі, обґрунтовує ефективність запропонованого методу.

У розділі 4 досліджено дві гібридні моделі, які ґрунтуються на моделях часових рядів. Перша модель являє собою неперервний та нелінійний аналог класичної лінійної ARMA моделі, друга модель являє собою узагальнення класичної LSTM нейронної мережі із врахуванням довільної скінченної передісторії. Перша модель базується на новому методі оцінки прогнозу процесів фінансових ринків в основі якого лежать моделі стохастичних диференціальних рівнянь із запізненням. Для оцінки параметрів в розробленій моделі використано генетичний алгоритм, а саме

метод рою частинок, оскільки цей метод виявився найбільш оптимальним в нашій задачі. Також вказано один із підходів переходу від оптимізаційної задачі по класу функцій до параметричної оптимізаційної задачі, що дозволяє значно ефективніше використовувати генетичні алгоритми пошуку оптимальних параметрів. Отримані теоретичні результати протестовано на реальних даних із використанням динаміки трьох реальних фінансових процесів. Результати тестування показали, що нова модель краще апроксимує реальні фінансові процеси, оскільки враховує як нелінійності моделей, так і нерегулярність між замірами реальних процесів. Друга модель розділу 4 даної роботи присвячена одному узагальненню LSTM нейронних мереж із врахуванням довжини передісторії нейронної мережі. Основна відмінність від класичної LSTM моделі або багатовимірного варіанту MV-LSTM моделі полягає у наявності гіперпараметру p , який відображає глибину залежності вхідного сигналу (x_t, \dots, x_{t+1-p}) та вихідного сигналу $g(x_{t+1})$ або в більш загальному випадку $g(x_{t+h})$ для $h > 0$.

У висновках підсумовано основні результати дисертаційного дослідження. **У додатках** подано наукові публікації, в яких відображено основні наукові результати роботи.

Теоретичне значення. Результати роботи, отримані в ході наукового дослідження, є удосконаленням теорії часових рядів та теорії машинного навчання. Підходи, запропоновані у дисертації, можуть використовуватися для подальших досліджень у цій галузі, у навчальних курсах кафедри прикладної математики та інформаційних технологій Чернівецького національного університету імені Юрія Федьковича (та інших ЗВО), пов'язаних з інтелектуальним аналізом даних, машинним навчанням, у методичних розробках, навчальних посібниках для освітнього процесу та науково-дослідної роботи студентів та аспірантів.

Практичне значення. Розроблені у дисертаційній роботі алгоритми можуть в подальшому використовуватися для практичного дослідження фінансових часових рядів. Основні результати роботи ґрунтуються на

твердженнях теорії ймовірності, статистики, хаотичних систем, теорії машинного навчання, випадкових процесів та теорії часових рядів.

Практичну цінність роботи проілюстровано на оцінюванні прогнозів реальних фінансових. Результати роботи знайдуть застосування у подальших дослідженнях із даної тематики.

Ключові слова: стохастичні диференціально-функціональні рівняння, генетичний алгоритм, прогнозування фінансових процесів, стохастична оптимізація, нейронні мережі, оцінка параметрів, метод зворотного поширення, часові ряди, штучний інтелект, машинне навчання, моделювання фондових індексів, часовий ряд фінансових даних, організаційна карта Кохонена, адаптивна нейро-нечітка система виводу, оцінка похибки.

ABSTRACT

Tokarieva K.A. Hybrid and ensemble machine learning methods and models for forecasting financial time series. – Qualification scientific work in the form of a manuscript.

Dissertation for the degree of Doctor of Philosophy in speciality 113 – Applied Mathematics. – Yuriy Fedkovych Chernivtsi National University, Ministry of Education and Science of Ukraine, Chernivtsi, 2024.

The dissertation is devoted to the construction of hybrid algorithms based on methods and models of machine learning and classical time series algorithms.

The results of the thesis are the basis for further theoretical and practical research developments in the study of hybrid forecasting algorithms for financial time series.

The thesis consists of an introduction, four chapters, conclusions, a list of references and an appendix.

The introduction substantiates the relevance of the research topic, formulates the purpose, objectives, subject, object and methods of the research, indicates the scientific novelty, theoretical and practical significance of the results obtained, analyses the relationship of the work with scientific research and the

personal contribution of the applicant, and provides information on the testing and publication of the main results of the dissertation. The structure and scope of the dissertation are described.

The first chapter reviews the scientific literature in the main areas of research on time series in general and financial time series in particular, considers key information on time series theory and machine learning methods, describes the main research areas and identifies the tasks that time series theory deals with, including the main problems of building a forecast for financial time series. A detailed overview of the main classical algorithms for building a forecast for a time series is provided, including their scope of application. The chronology of the development of time series theory and conditional periods of model complexity are analysed in detail. The second section of this chapter discusses the main models of artificial intelligence and machine learning for time series analysis, describes the topologies of neural networks that will be used in the construction of hybrid algorithms.

In the second chapter, the author investigate hybrid models in which the main process is estimated on the basis of a time series model. Model residuals are estimated on the basis of neural networks. The use of this approach allowed us to extend the use of classical time series models to the case of the absence of homoscedasticity in the residuals and the absence of non-smoothness of the time series residuals. The main results are:

- time series of stock indices contain both linear and non-linear components, and therefore, a linear ARIMA model and a non-linear ANN model alone cannot provide an accurate estimate of such time series;
- a hybrid model of an autoregressive moving average model and an artificial neural network for forecasting time series of stock indices is developed;
- it is proved that if the time series is strictly stationary, the ARIMA model is more adequate for Gaussian series in the developed hybrid model;
- a comparative analysis of the proposed algorithm with classical algorithms for estimating the forecast of financial time series is carried out, as a result

of which the developed method demonstrates greater accuracy compared to other hybrid models, since the data series is decomposed into components with low and high volatility (linear and nonlinear), and therefore the total error of the model is low.

The third chapter considers the relationship between chaotic dynamical systems and time series systems. The main approaches to the definition of deterministic chaos and the main methods of taking into account the presence of deterministic chaos when assessing the forecast of a financial time series are analysed. Four methods are presented to determine the existence of chaos in the system:

- Fourier transform;
- Lyapunov exponent;
- Hurst exponent;
- fractal dimension.

The main approach to the application of chaos theory, namely deterministic chaos, is to use the Tuckers embedding theorem, according to which the structure of a chaotic deterministic process can be simplified by using a partition of the vector space $(x(t), \dots, x(t + T))$. The phase space reconstruction method is used to reconstruct phase space points from chaotic time series using the embedding dimension D and time delay T . According to the phase space reconstruction method, the phase space points reproduce the key characteristics of the original time series. In order to confirm the effectiveness of the proposed methodology, three different classical chaotic time series were studied in this section: McKay-Glass, logistic and Eno. The proposed method was combined with genetic algorithms to improve the forecasting results. It was shown that the new forecasting method improved with GA demonstrates outstanding results. Comparison of the results of the proposed method with the results reported in the literature substantiates the effectiveness of the proposed method.

In chapter four, the author examines two hybrid models based on time series models. The first model is a continuous and nonlinear analogue of the

classical linear ARMA model, while the second model is a generalisation of the classical LSTM neural network with an arbitrary finite backwardation. The first model is based on a new method of estimating the forecast of financial market processes based on stochastic differential equation models with a delay. A genetic algorithm, namely the particle swarm method, was used to estimate the parameters in the developed model, as this method proved to be the most optimal for our problem. The article also presents one of the approaches to the transition from a function class optimisation problem to a parametric optimisation problem, which allows a much more efficient use of genetic algorithms to find optimal parameters. The theoretical results obtained are tested on real data using the dynamics of three real financial processes. The test results show that the new model better approximates real financial processes, as it takes into account both the non-linearity of the models and the irregularity between measurements of real processes. The second model in chapter 4 of this paper is dedicated to a generalisation of LSTM neural networks, taking into account the length of the back-propagation of the neural network. The main difference with the classical LSTM model or the multivariate variant of the MV-LSTM model is the presence of the hyperparameter p , which reflects the depth of dependence of the input signal (x_t, \dots, x_{t+1-p}) and output signal $g(x_{t+1})$ or in a more general case $g(x_{t+h})$ для $h > 0$.

The **conclusions** summarise the main findings of the thesis research, and the **appendices** contain scientific publications that reflect the main scientific findings of the thesis.

Theoretical significance. The results obtained in the course of the research improve the theory of time series and machine learning. The approaches proposed in the thesis can be used for further research in this area, in the courses of the Department of Applied Mathematics and Information Technologies of the Yuriy Fedkovych Chernivtsi National University (and other universities) related to data mining, machine learning, in methodological developments, textbooks for the educational process and research work of students and postgraduates.

Practical significance. The algorithms developed in this thesis can be further used for practical research on financial time series. The main results of the thesis are based on probability theory, statistics, chaotic systems, machine learning theory, random processes and time series theory.

The practical value of the work is illustrated by evaluating forecasts of real financial markets. The results of the work will be used for further research on this topic.

Keywords: stochastic differential functional equations, genetic algorithm, financial forecasting, stochastic optimisation, neural networks, parameter estimation, back-propagation, time series, artificial intelligence, machine learning, stock index modelling, financial time series, Kohonen chart, adaptive neuro-fuzzy inference system, error estimation.

СПИСОК ПУБЛІКАЦІЙ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Наукові праці у періодичних наукових виданнях, проіндексованих у наукометричній базі даних Scopus:

1. Kushnir, M., Tokarieva, K. A Generalization of the Arima Model to the Nonlinear and Continuous Cases. *Cybernetics and Systems Analysis*. 2023. Vol. 59. Iss. 6. P. 900–909 (Scopus), (Q3 – <https://link.springer.com/article/10.1007/s10559-023-00625-8>)
2. Кушнір М.Я., Токарева К.А. Використання Систем Штучного Інтелекту У Задачах Прогнозування Фінансових Індексів: Огляд Наукових Джерел / Artificial Intelligence Systems In The Financial Market Predictions: Literature Review. *Radioelectronic and Computer Systems*. 2020. Is. 3(95). С. 108-117 (Scopus), (URL: <https://www.scopus.com/record/display.uri?eid=2-s2.0-85101514809&origin=resultslist>)

Наукові праці у виданнях, включених до переліку наукових фахових видань України:

3. Кушнір М.Я., Токарева К.А. Одне узагальнення ARIMA-моделі на нелінійний та неперервний випадки / A Generalization of the Arima Model to the Nonlinear and Continuous Cases. *Cybernetics and Systems Analysis*. 2023. Vol. 59. Iss. 6. P. 900 – 909. (Scopus), (Q3 - URL: <https://www.scimagojr.com/journalsearch.php?q=12933&tip=sid&clean=0>)
4. Кушнір М.Я., Токарева К.А. Гібридна модель самоорганізаційної карти Кохонена та адаптивної нейро-нечіткої системи виводу у задачах прогнозування цін фондових індексів. *Буковинський математичний журнал*. 2021. Т. 9. № 2. С. 70–80. (URL: <https://doi.org/10.31861/bmj2021.02.05>)
5. Кушнір, М. Я., Токарева, К. А. Одне узагальнення LSTM-нейронних мереж. *International Scientific Technical Journal "Problems of Control and Informatics"*. 2023. 68(4). С. 111–118. (URL: <https://jais.net.ua/index.php/files/article/view/118>)

Наукові праці, які засвідчують апробацію матеріалів дисертації:

6. M. Kushnir, A. Komarnytskyi, K. Tokarieva, N. Savchyn, P. Kroialo and V. Toronchuk, Technological and Legal Aspects of the Use of Machine Learning Elements in Chaotic Information Processing Systems. *2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T)*, Kharkiv, Ukraine, 6-9 October 2020, P. 713-716. ISBN: 978-172819177-5 (Scopus) (URL: <https://www.scopus.com/record/display.uri?eid=2-s2.0-85114410889&origin=resultslist>)
7. V. Hryhorkiv, L. Buiak, A. Verstiak, M. Hryhorkiv, O. Verstiak and K. Tokarieva, "Forecasting Financial Time Series Using Combined ARIMA-ANN Algorithm," *2020 10th International Conference on Advanced Computer Information Technologies (ACIT)*, Deggendorf, Germany, 16-18 September 2020, P. 455-458. ISBN: 978-172816760-2 (Scopus), (URL: <https://www.scopus.com/record/display.uri?eid=2-s2.0-85094125855&origin=resultslist>)
8. Токарева К.А. Використання систем штучного інтелекту у прогнозуванні фінансових ринків. *Міжнародний науковий симпозиум «Інтелектуальні рішення»*. Обчислювальний інтелект (результати, проблеми, перспективи): праці міжнар. наук.-практ. конф., 15-20 квітня 2019 р., Ужгород / М-во освіти і науки України, ДВНЗ «Ужгородський національний університет», та [ін.]; наук. ред. В.Є. Снитюк. С. 54-55. (URL: <https://er.chdtu.edu.ua/bitstream/ChSTU/3192/1/OI-2019.pdf>)
9. Токарева К.А. Лог-рентабельність активів у дослідженні фінансових часових рядів. *Математичні методи, моделі та інформаційні технології в економіці*: Матеріали VI Міжнародної науково-методичної конференції. Чернівці: Друк Арт, 6 вересня 2019. С. 154-155.