# Облікова картка дисертації

## I. Загальні відомості

**Державний обліковий номер:** 0823U101894

**Особливі позначки:** відкрита

**Дата реєстрації:** 19-12-2023

**Статус:** Наказ про видачу диплома

**Реквізити наказу МОН / наказу закладу:** Міністерство освіти і науки України ЧЕРНІВЕЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені Юрія Федьковича НАКАЗ від 8 січня 2024 року №7 про видачу диплома доктора філософії Кнігніцькій Тетяні Василівні

## II. Відомості про здобувача

**Власне Прізвище Ім'я По-батькові:**

1. Кнігніцька Тетяна Василівна
2. Knignitska Tetyana V

**Кваліфікація:**

**Ідентифікатор ORCID ID:** Не застосовується

**Вид дисертації:** доктор філософії

**Аспірантура/Докторантура:** так

**Шифр наукової спеціальності:** 113

**Назва наукової спеціальності:** Прикладна математика

**Галузь / галузі знань:** математика та статистика

**Освітньо-наукова програма зі спеціальності:** 38612 Прикладна математика (113 Прикладна математика)

**Дата захисту:** 22-12-2023

**Спеціальність за освітою:** Статистика

**Місце роботи здобувача:**

**Код за ЄДРПОУ:**

**Місцезнаходження:**

**Форма власності:**

**Сфера управління:**

**Ідентифікатор ROR:** Не застосовується

## III. Відомості про організацію, де відбувся захист

**Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради):** ДФ 76.051.037 (ID 2714)

**Повне найменування юридичної особи:** Чернівецький національний університет імені Юрія Федьковича

**Код за ЄДРПОУ:** 02071240

**Місцезнаходження:** вул. Коцюбинського, буд. 2, Чернівці, 58012, Україна

**Форма власності:** Державна

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:**

## IV. Відомості про підприємство, установу, організацію, в якій було виконано дисертацію

**Повне найменування юридичної особи:** Чернівецький національний університет імені Юрія Федьковича

**Код за ЄДРПОУ:** 02071240

**Місцезнаходження:** вул. Коцюбинського, буд. 2, Чернівці, 58012, Україна

**Форма власності:** Державна

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:**

## V. Відомості про дисертацію

**Мова дисертації:** Українська

**Коди тематичних рубрик:** 27.43.17, 83.77.23, 83.77.29, 83.77.31, 27.43.15

### Тема дисертації:

1. Оцінки параметрів авторегресійних моделей
2. Autoregressive models parameters estimations

### Реферат:

1. Дисертаційна робота присвячена знаходженню відстаней між вимірюваннями даних, які представлені часовими рядами, та визначенню оптимальної кількості кластерів на основі власних значень стохастичної матриці графа. Дисертація складається із вступу, трьох розділів, висновків та переліку використаних джерел. У вступі обгрунтовано актуальність теми дослідження, сформульовано мету, завдання, предмет, об'єкт та методи дослідження, вказано наукову новизну, практичне значення отриманих результатів, зв'язок роботи з науковими дослідженнями та особистий внесок здобувача, а також наведено дані про те, де доповідались, обговорювались та були опубліковані основні результати дисертації. У першому розділі здійснено огляд наукової літератури, присвяченої дослідженню часових рядів, зокрема, визначенню метрик подібності між часовими рядами та підходи до кластеризації даних, які представлені у вигляді неструктурованих типів

даних. Детально проаналізовано хронологію розвитку наукових підходів до задач кластеризації, класифікації, зменшення розмірності часових рядів. Розділ 1 відображає загальний огляд розвитку наукових досліджень при дослідженні часових рядів та існуючі метрики для встановлення подібності між часовими рядами. Тут наведено методи дослідження структурних стрибків у часових рядах та зроблено огляд наукових досліджень, які стосуються неперервних часових рядів. Вибір оптимальної кількості кластерів при поділі даних на групи також представлено у розділі 1. У другому розділі запропоновано визначати подібність або відстань між часовими рядами за допомогою моделей часових рядів. Запропонований алгоритм для встановлення подібності двох наборів даних використовує параметри моделі, а не самі вимірювання. У якості моделей часових рядів розглянуто стаціонарні ARMA моделі. Отриманий алгоритм порівнюється з уже існуючими метриками знаходження відстаней у випадку збільшення вимірювань часового ряду та у випадку зростання кількості викидів у вхідному часовому ряді. Отриманий алгоритм має меншу обчислювальну складність, ніж алгоритми Евкліда, DTW та ERP. Запропоновану відстань можна використовувати для кластеризації сильно зашумлених даних. Наукову новизну висновків, зроблених на основі отриманих у другому розділі результатів, розкривають такі положення: Описано алгоритм для знаходження відстані між часовими рядами на основі моделей часових рядів. Отримана відстань є більш стійкою до викидів у часових рядах. У випадку збільшення кількості викидів запропонований у дисертаційному дослідженні алгоритм дає кращі результати (відносна похибка зростає логарифмічно), ніж аналогічні алгоритми (Евклідова відстань, ERP, DTW) для знаходження відстані між часовими рядами (відносна похибка зростає лінійно). Запропонований метод знаходження відстані між вимірюваннями часового ряду дає кращі результати для великих часових рядів, коли кількість вимірювань T > 1000. До того ж обчислювальна складність отриманого алгоритму є меншою за обчислювальну складність уже існуючих алгоритмів. У третьому розділі розглянуто проблему кластеризації на графах на основі власних значень стохастичної матриці графа. Доведено, що власні значення стохастичної матриці для великих графів (N >100) поділяються на три групи, одна із яких є визначальною для числа кластерів у графі. Використовуючи теорію випадкових матриць, вдалося показати, що асимптотичний розподіл підгрупи дійсних частин власних значень стохастичної матриці графу описується напівколовим розподілом Вігнера. Використання стохастичних матриць дало змогу точно локалізувати власні значення, що відповідають за кількість кластерів, чого не вдавалося зробити для матриць суміжності. Основні припущення моделі пов'язані з властивостями дискретних ланцюгів Маркова, що дозволяє розширити область застосування отриманих результатів на більш широкий клас об'єктів. Теоретичні результати перевірені на кластеризації часових рядів,що описують вартості N = 470 акцій S&P500 в період з 2013 до 2018 року. Наукову новизну висновків, зроблених на основі отриманих у третьому розділі результатів, розкривають такі положення: У роботі запропоновано новий метод визначення оптимальної кількості кластерів при кластеризації об'єктів, що задаються неструктурованими даними (графами та часовими рядами) на основі спектрального аналізу стохастичної матриці даного графу. Використовуючи метод Монте-Карло, вдалося показати, що запропонований метод дає кращі результати для визначення оптимальної кількості кластерів у порівнянні із деякими класичними методами.

2. The dissertation work is devoted to finding the distances between data measurements, which are represented by time series, and determining the optimal number of clusters based on the eigenvalues of the stochastic matrix of the graph. The dissertation consists of an introduction, three sections, conclusions and a list of used sources. The introduction substantiates the relevance of the research topic, formulates the goal, task, subject, object and research methods, indicates the scientific novelty, the practical significance of the results obtained, the connection of the work with scientific research and the personal contribution of the recipient, and also provides data on where the main results of the dissertation were reported, discussed and published. In the first section, a review of the scientific literature devoted to the study of time series, in particular, the determination of similarity metrics between time series and approaches to clustering data, which are presented in the form of unstructured data types, is carried out. The chronology of the development of scientific approaches to the problems of clustering, classification, and dimensionality reduction of time series is analyzed in detail. Chapter 1 presents a general overview of the development of scientific research in the study of time series and existing metrics for

establishing similarity between time series. Methods for studying structural jumps in time series are presented here, and an overview of scientific research related to continuous time series is provided. The selection of the optimal number of clusters when dividing the data into groups is also presented in section 1. In the second section, it is proposed to determine the similarity or distance between time series using time series models. The proposed algorithm uses model parameters rather than the measurements themselves to establish the similarity between two data sets. Stationary ARMA models are considered as time series models. The resulting algorithm is compared with already existing metrics for finding distances in the case of an increase in time series measurements and in the case of an increase in the number of outliers in the input time series. The resulting algorithm has lower computational complexity than the Euclidean, DTW, and ERP algorithms. The proposed distance can be used for clustering highly noisy data. The scientific novelty of the conclusions drawn on the basis of the results obtained in the second section is revealed by the following provisions: An algorithm for finding the distance between time series based on time series models is described. The resulting distance is more robust to outliers in the time series. In the case of an increase in the number of emissions, the algorithm proposed in the dissertation research gives better results (the relative error increases logarithmically) than similar algorithms (Euclidean distance, ERP, DTW) for finding the distance between time series (the relative error increases linearly). The proposed method of finding the distance between time series measurements gives better results for large time series when the number of measurements T > 1000. In addition, the computational complexity of the obtained algorithm is lower than the computational complexity of already existing algorithms. In the third section, the problem of clustering on graphs based on the eigenvalues of the stochastic matrix of the graph is considered. It is proved that the eigenvalues of the stochastic matrix for large graphs (N >100) are divided into three groups, one of which is the determinant for the number of clusters in the graph. Using the theory of random matrices, it was possible to show that the asymptotic distribution of the subgroup of the real parts of the eigenvalues of the stochastic matrix of the graph is described by the semicircular Wigner distribution. The use of stochastic matrices made it possible to precisely localize the eigenvalues responsible for the number of clusters, which could not be done for adjacency matrices. The main assumptions of the model are related to the properties of discrete Markov chains, which makes it possible to expand the scope of the obtained results to a wider class of objects. The theoretical results were tested on the clustering of time series describing the values of N = 470 S&P500 shares in the period from 2013 to 2018. The scientific novelty of the conclusions drawn on the basis of the results obtained in the third section is revealed by the following provisions: The paper proposes a new method for determining the optimal number of clusters when clustering objects given by unstructured data (graphs and time series) based on the spectral analysis of the stochastic matrix of the given graph. Using the Monte Carlo method, it was possible to show that the proposed method gives better results for determining the optimal number of clusters in comparison with some classical methods.

**Державний реєстраційний номер ДіР:** 0102U006591

**Пріоритетний напрям розвитку науки і техніки:** Фундаментальні наукові дослідження з найбільш важливих проблем розвитку науково-технічного, соціально-економічного, суспільно-політичного, людського потенціалу для забезпечення конкурентоспроможності України у світі та сталого розвитку суспільства і держави

**Стратегічний пріоритетний напрям інноваційної діяльності:** Розвиток сучасних інформаційних, комунікаційних технологій, робототехніки

**Підсумки дослідження:** Нове вирішення актуального наукового завдання

**Публікації:**

- 1. Knignitskaya T. V. Estimate of time series similarity based on models. Journal of Automation and Information Sciences. 2019. Vol. 51 (№8). (Scopus) (Q3 – https://www.scimagojr.com/journalsearch.php?q=25497&tip=sid&clean=0) – 2 публікації DOI:

10.1615/JAutomatInfScien.v51.i8.60 https://www.scopus.com/record/display.uri?eid=2-s2.0-85077811133&origin=resultslist&sort=plf-f&src=s&sid=3961a88e8343ba78b64b63dfe0112e28&sot=b&sdt=b&s=TITLE%28Estimate+of+Time+Series+Similarity+B

- 2. Pavlyukovich N., Pavlyukovich O.V., Dubolazov O.V., Ushenko Yu.A., Tomka Yu. Ya., Zabolotna N.I., Soltys I.V., Drin Ya.M., Knignitska T.V., Talakh M.V., Dovgun A.Ya., Kotyra A., and Kozbakova A. Methods and means of "single-point" phasometry of microscopic images of optical-anisotropic biological objects. Proceedings of SPIE - The International Society for Optical Engineering. Vol. 11176, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2019, 1117630 (6 November 2019). ISSN: 0277786X (Scopus) – 1 публікація DOI: 10.1117/12.2537168 https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11176/1117630/Methods-and-means-of-single-point-phasometry-of-microscopic-images/10.1117/12.2537168.short
- 3. Кнігніцька Т.В., Малик І.В., Горбатенко М.Ю. Кластеризація: марковський алгоритм Буковинський математичний журнал. 2020. 7(2). С. 59-75. (фахова) – 0,5 публікації DOI: 10.31861/bmj2019.02.059 http://bmj.fmi.org.ua/index.php/adm/article/view/944

**Наукова (науково-технічна) продукція:** технології

**Соціально-економічна спрямованість:** збільшення обсягів виробництва; поліпшення стану навколишнього середовища; економія енергоресурсів; підвищення продуктивності праці; поліпшення якості життя та здоров'я населення, ефективності діагностики та лікування хворих; підвищення автоматизації виробничих процесів

**Охоронні документи на ОПІВ:**

Комп'ютерні програми

Программа для визначення оптимальної кількості кластерів та встановлення подібності

**Впровадження результатів дисертації:** Планується до впровадження

**Зв'язок з науковими темами:**

## VI. Відомості про наукового керівника/керівників (консультанта)

**Власне Прізвище Ім'я По-батькові:**

1. Малик Ігор Володимирович

2. Igor V. Malyk

**Кваліфікація:** д. ф.-м. н., доц., 01.05.01

**Ідентифікатор ORCID ID:** Не застосовується

**Додаткова інформація:**

**Повне найменування юридичної особи:** Чернівецький національний університет імені Юрія Федьковича

**Код за ЄДРПОУ:** 02071240

**Місцезнаходження:** вул. Коцюбинського, буд. 2, Чернівці, 58012, Україна

**Форма власності:** Державна

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:**

## VII. Відомості про офіційних опонентів та рецензентів

### Офіційні опоненти

**Власне Прізвище Ім'я По-батькові:**

1. Самойленко Ігор Валерійович
2. Igor V. Samoilenko

**Кваліфікація:** д.ф.-м.н., професор, 01.05.04

**Ідентифікатор ORCID ID:** Не застосовується

**Додаткова інформація:**

**Повне найменування юридичної особи:** Київський національний університет імені Тараса Шевченка

**Код за ЄДРПОУ:** 02070944

**Місцезнаходження:** вул. Володимирська, буд. 60, Київ, 01033, Україна

**Форма власності:** Державна

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:**

**Власне Прізвище Ім'я По-батькові:**

1. Кнопов Павло Соломонович
2. Pavlo S. Knopov

**Кваліфікація:** д.ф.-м.н., професор, 01.01.09

**Ідентифікатор ORCID ID:** Не застосовується

**Додаткова інформація:**

**Повне найменування юридичної особи:** Інститут кібернетики імені В. М. Глушкова Національної академії наук України

**Код за ЄДРПОУ:** 05417176

**Місцезнаходження:** проспект Академіка Глушкова, буд. 40, Київ, 03187, Україна

**Форма власності:** Державна

**Сфера управління:** Національна академія наук України

**Ідентифікатор ROR:**

### Рецензенти

**Власне Прізвище Ім'я По-батькові:**

1. Угрин Дмитро Ілліч
2. Dmytro I. Ugryn

**Кваліфікація:** д. т. н., доц., 05.13.06

**Ідентифікатор ORCID ID:** Не застосовується

**Додаткова інформація:**

**Повне найменування юридичної особи:** Чернівецький національний університет імені Юрія Федьковича

**Код за ЄДРПОУ:** 02071240

**Місцезнаходження:** вул. Коцюбинського, буд. 2, Чернівці, 58012, Україна

**Форма власності:** Державна

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:**

**Власне Прізвище Ім'я По-батькові:**

1. Юрченко Ігор Валерійович
2. Ihor V. Yurchenko

**Кваліфікація:** к. ф.-м. н., доц., 01.05.02

**Ідентифікатор ORCID ID:** Не застосовується

**Додаткова інформація:** ;https://www.scopus.com/authid/detail.uri?authorId=23096632000

**Повне найменування юридичної особи:** Чернівецький національний університет імені Юрія Федьковича

**Код за ЄДРПОУ:** 02071240

**Місцезнаходження:** вул. Коцюбинського, буд. 2, Чернівці, 58012, Україна

**Форма власності:** Державна

**Сфера управління:** Міністерство освіти і науки України

**Ідентифікатор ROR:**

## VIII. Заключні відомості

| | |
|---|---|
| **Власне Прізвище Ім'я По-батькові голови ради** | Григорків Василь Степанович |
| **Власне Прізвище Ім'я По-батькові головуючого на засіданні** | Григорків Василь Степанович |
| **Відповідальний за підготовку облікових документів** | Кнігніцька Тетяна Василівна |
| **Реєстратор** | УкрІНТЕІ |

**Керівник відділу УкрІНТЕІ, що є відповідальним за реєстрацію наукової діяльності**

Юрченко Тетяна Анатоліївна