

Н. О. Чікіна, І. В. Антонова

Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

ПОРІВНЯЛЬНИЙ АНАЛІЗ ЧАСОВИХ РЯДІВ

Анотація. Актуальність. Статистичні дані по більшості соціальних, демографічних, медичних, екологічних показниках збираються частіше за все раз на рік. Такі ряди доступні лише за невеликими за обсягом даними в декілька років, тобто є короткими часовими рядами. Більшість існуючих методів кластеризації не є придатними для аналізу коротких часових рядів. Тому актуальною є задача подання інформації про числові ряди таким чином, щоб вона була достатньою для цілей задач кластеризації саме для коротких часових рядів. Останнім часом в науковій літературі з'явилися публікації з рекомендаціями надавати паралельно додаткову інформацію, що міститься у статистичних показниках вимірювань часових рядів, а також показники, які описують динаміку часових рядів, їх мінливість, але не залежать від часу. Залишається актуальною задача вибору міри близькості рядів. Існуючі дослідження щодо порівняльного аналізу міри близькості схиляються до того, що Евклідова відстань займає перші місця за якістю в задачах кластеризації часових рядів. **Об'єкт дослідження:** короткі часові ряди даних розповсюдженості окремих видів шкірних захворювань в регіонах України. **Мета статті:** порівняльний аналіз результатів кластеризації коротких часових рядів, отриманих із застосуванням різних мір близькості. **Результати дослідження.** У роботі досліджено часові ряди, що відображають дані з кількості випадків захворювань різними шкірними хворобами в регіонах України за певний період часу. Результати кластеризації таких часових рядів дають можливість зробити порівняльний аналіз даних і виявити групи регіонів України з «однаковою» ситуацією з цього питання. **Висновки.** Отримані результати можуть бути використані в подальших дослідженнях для виявлення причин відповідних рівнів захворюваності, таких, наприклад, як якість води, повітря, забруднення навколишнього середовища. Для вирішення цієї задачі автори, крім міри Евклідової відстані, застосовують міру близькості часових рядів за їх основними статистичними характеристиками та міру близькості показників динаміки досліджуваних рядів, а саме, рядів їх перших різниць. З методів кластеризації часових рядів обрана ієрархічна процедура. Проведено порівняльний аналіз отриманих результатів.

Ключові слова: математичне моделювання; короткий часовий ряд; порівняльний аналіз; ієрархічна процедура кластерного аналізу; міри близькості; дендрограма; функціонал якості.

Вступ

Постановка проблеми. Останнім часом спостерігається сплеск появи даних, що представлені у вигляді часових рядів. Це спровокувало зріст наукових досліджень з ціллю виявлення з цих даних різноманітної інформації. Основною задачею аналізу часових рядів залишається задача прогнозування його подальших значень. Одночасно з'являються наукові публікації щодо їх порівняльного аналізу за допомогою різних мір близькості.

Кластерний аналіз відноситься до класу методів порівняльного аналізу даних, який об'єднує методи класифікації, що засновані на визначенні поняття близькості (відстані) між об'єктами з подальшим виділенням з них груп (кластерів).

Мета кластеризації полягає в тому, щоб отримати такі групи (кластери), в яких відмінності між ознаками об'єктів всередині цих груп були б мінімальні, а відмінності між об'єктами – максимальні.

Кластерний аналіз не вимагає апріорних відомостей про розподіл даних. Часові ряди, на відміну від статичних даних, складаються з величин, що характеризують об'єкти дослідження, стан яких змінюється з часом. Тому використання кластерного аналізу для їх дослідження є можливим.

Існує достатньо велика кількість алгоритмів кластеризації. Однак, згідно з визначенням, наведеним в [1], всі методи кластеризації можна умовно поділити тільки на два класи: ієрархічні та засновані на поділі даних. В цілому, методи, засновані на поділі даних, в даний час є найбільш популярними, багато в чому завдяки їх математичній строгості, ясності

одержуваних результатів і просунутому програмному забезпеченню.

Робота кластерного аналізу спирається на два припущення. По-перше, це те, що об'єкти дослідження в принципі допускають розбиття сукупності на кластери. По-друге, правильність вибору масштабу або одиниці вимірювання ознак. Вибір масштабу в кластерному аналізі має велике значення. Якщо вимірювання часового ряду на порядок більші відповідних вимірювань ряду, то при розрахунках величини близькості між рядами, що відображають положення об'єктів у просторі їх властивостей, це стає неможливим коректно виконати. Зазначена проблема вирішується за допомогою попередньої стандартизації змінних. Стандартизація (standardization) або нормування (normalization) приводить значення всіх перетворених змінних до єдиного діапазону значень шляхом вираження через відношення цих значень до якоїсь величини, що відображає певні властивості конкретної ознаки.

Якість кластеризації часових рядів, як і будь-яких інших об'єктів дослідження за цим методом, залежить, по-перше, від обраного алгоритму кластеризації і, по-друге, від обраної міри близькості між рядами, що є характеристикою ступеня різниці між ними. В подібних задачах обрання міри близькості впливає на якість в більшій мірі порівняно з обранням методу кластеризації [2]. Це, в свою чергу, викликало появу досліджень стосовно ефективності міри близькості часових рядів для вирішення задач кластеризації [3, 4].

Аналіз останніх досліджень. Кластерний аналіз часових рядів використовується для вирішення

багатьох практичних задач. Наприклад, аналіз динаміки захворюваності або станів пацієнтів [5–8], фінансових показників [9], в задачах прогнозування [10], аналізу ризиків [11] тощо.

Як показує аналіз літератури, на практиці при вивченні часових рядів у більшості робіт використовується ієрархічна кластеризація [5, 6, 9], метод кластеризації K - середніх [10, 11], а також нечітка кластеризація K - середніх [12], при якій кластери можуть перетинатися.

Важливу роль в задачах кластеризації відіграють ієрархічні процедури. Ієрархічна процедура починається з того, що кожен об'єкт є окремим кластером. Потім два сусідніх кластера об'єднуються в один, і так далі, і цей процес триває до тих пір, поки не залишаться два кластери.

Метод K - середніх є також одним з найбільш застосованих методів кластеризації [8, 13]. Процедура починається з визначення кількості кластерів і завдання центрів майбутніх кластерів – таких об'єктів, в околі яких формуються кластери. На наступному етапі об'єкти групуються відповідно до їх близькості до центрів кластерів. Процедура кластеризації виконується до тих пір, поки центри не перестануть змінюватись.

До недоліків методу K - середніх можна віднести те, що кількість кластерів не змінюється, а результат залежить від початкового визначення центрів.

Якщо досліджуються короткі ряди, то більшість існуючих методів кластеризації не призначені для їх аналізу.

Що стосується обрання міри близькості між рядами, що є характеристикою ступеня різниці між ними, це є також важливим моментом проведення процедури кластеризації.

Існує три підходи до розрахунку міри близькості між часовими рядами [13]:

1. Підхід, заснований на необроблених даних. Вимірювання відстані обчислюються з використанням вихідних значень аналізованих часових рядів або деяких їхніх характеристик. Найбільш відомі міри відстані цього підходу: Евклідова, Мінковського, Манхеттенська відстані, відстань між періодограмами, відстань між коефіцієнтами автокореляцій рядів та інші.

2. Підхід, заснований на наближеннях часових рядів моделями ARIMA. У цьому підході розрахунок близькості між окремими рядами виконується за значеннями статистичних характеристик моделей. При застосуванні цього підходу будується статистика, що відображає значущість відмінностей процесів, наприклад, коефіцієнти авторегресії або матриці коваріацій рядів та інше.

3. В рамках цього підходу оцінюється рівень інформації, що міститься у кожному з часових рядів. Цей підхід не враховує часові характеристики і особливості процесів, не робить припущень про моделі, які можуть бути використані для апроксимації процесу.

Кожний з перелічених підходів має свою специфіку. Обрання методу дослідження повинне

відповідати цілям задачі, що вирішується.

Оскільки процедура кластерного аналізу проходить в автоматичному режимі без участі дослідників, то дуже складно зробити об'єктивний висновок про ступінь «правильності» отриманих з його допомогою результатів. Як правило, на практиці обирається той результат, який «має сенс» з точки зору розв'язуваної задачі. Однак є низка показників, які можна використовувати для кількісного опису якості кластеризації. Це дає можливість порівняти різні рішення та обрати найбільш «оптимальне». У зв'язку з цим розрахунок таких мір якості зазвичай є частиною алгоритму кластерного аналізу.

Для визначення того, яка кількість класів є оптимальною, вирішальне значення має показник, який обчислює близькість між двома кластерами згідно обраної міри. На тому етапі, де ця міра близькості між двома кластерами збільшується перервами, процес об'єднання в нові кластери необхідно зупинити, тому що були б вже об'єднані кластери, які знаходяться на відносно великій відстані один від одного.

Ступінь наближення результатів до ідеального виконання процедури кластеризації характеризується також функціоналом якості $\frac{F_0}{F_1}$. Для його побудови

обчислюють суму F_0 внутрішньокластерних відстаней між об'єктами в кожному кластері та суму F_1 міжкластерних відстаней між об'єктами, які знаходяться в різних кластерах. У цьому випадку якість розбиття на кластери за допомогою сум F_0 і F_1 оцінюється співвідношенням $\frac{F_0}{F_1}$.

Вважається, що найкращому поділу множини $\Omega = \{X_1, X_2, \dots, X_m\}$ на кластери відповідає мінімальне значення співвідношення $\frac{F_0}{F_1}$.

Метою роботи є порівняльний аналіз часових рядів даних розповсюдженості окремих видів шкірних захворювань в регіонах України за результатами кластерного аналізу із застосуванням різних мір близькості.

Основний матеріал

Нехай досліджується сукупність $\Omega = \{X_1, X_2, \dots, X_m\}$ з m коротких часових рядів з однаковою в часі кількістю вимірювань n . Треба поділити сукупність Ω на групи (кластери) C_1, C_2, \dots, C_k , де $\Omega = C_1 \cup C_2 \cup \dots \cup C_k$, $C_i \cap C_j = \emptyset$, $i, j = \overline{1, k}$, $i \neq j$, застосовуючи обрану процедуру ієрархічної кластеризації та міру $D_q(X_i, X_j)$ ($q = 1, 2, 3$) близькості між рядами X_i, X_j ($i, j = \overline{1, m}$).

Математична модель. Нехай задана скінченна множина об'єктів $\Omega = \{X_1, X_2, \dots, X_m\}$. Множиною \mathcal{C} кластерів цих об'єктів називають будь-яку сукупність його непустих підмножин, що не перети-

наються. Таким чином, якщо $C \subseteq 2^\Omega$ – множина кластерів, то

$$\forall C_1, C_2 \in C: (C_1 \cap C_2 \neq \emptyset) \Leftrightarrow (C_1 = C_2).$$

Для вирішення поставленої задачі пропонується провести ієрархічну процедуру (IP) кластерного аналізу, що визнана найбільш застосованою в дослідницьких цілях.

Доведено, що результати кластеризації певним чином залежать від обраної міри близькості між об'єктами – рядами. При проведенні ієрархічної процедури (IP) застосуємо три міри близькості: $D_1(X_i, X_j)$ – Евклідову міру, якою залежність вимірювань від часу ігнорується; $D_2(ST_i, ST_j)$ – міру, що побудована на статистичних ST -характеристиках часових рядів, для яких не враховується залежність від часу; та $D_3(\Delta X_i, \Delta X_j)$ – міру, що побудована на перших різницях

$$\Delta x_i(x_{i,r}, x_{i,r+1}) = x_{i,r+1} - x_{i,r} \quad (i = \overline{1, m}, r = \overline{1, n-1})$$

вимірювань даних часових рядів. Таким чином, буде отримано три варіанта вирішення задачі кластеризації множини $\Omega = \{X_1, X_2, \dots, X_m\}$ часових рядів із застосуванням ієрархічної процедури за обраними мірами близькості (табл. 1).

Аналіз отриманих результатів може дати можливість виявити переваги (недоліки) застосування мір близькості типу $D_2(ST_i, ST_j)$ та $D_3(\Delta X_i, \Delta X_j)$ перед найбільш застосованою у дослідженнях мірою відстані $D_1(X_i, X_j)$.

Результати досліджень. В якості об'єкта $\Omega = \{X_1, X_2, \dots, X_m\}$ розглянемо $m = 3$ регіонів України, кожен з яких характеризується відповідним часовим рядом X_i ($i = \overline{1, m}$) з однаковою в часі кількістю $n = 13$ вимірювань випадків деяких шкірних захворювань [14]. Для дослідження обрані три з п'яти регіонів України, а саме Східний, Центральний і Західний. Дані розповсюдження захворювань по регіонах представлені на рис. 1.

Поділ часових рядів на кластери проводився ієрархічна процедурою (Hierarchical Cluster) методом Between-groups Linkage (зв'язок між групами) з використанням математичного пакету SPSS Statistics 26.

Перед проведенням кластерного аналізу виконаємо нормування вимірювань часових рядів. Використання L_1 -норми для часових рядів можливе лише у випадку, коли ряди однакової довжини й не мають локального зсуву за часом.

При проведенні ієрархічної процедури (IP) було застосовано три міри близькості:

1) евклідова міра:

$$D_1(X_i, X_j) = \sqrt{\sum_{l=1}^n (x_{il} - x_{jl})^2};$$

2) міра, що побудована на p статистичних ST -характеристиках s_{il} ($i = \overline{1, m}, l = \overline{1, p}$) нормованих вимірювань часових рядів:

$$D_2(ST_i, ST_j) = \sqrt{\sum_{l=1}^p (s_{il} - s_{jl})^2};$$

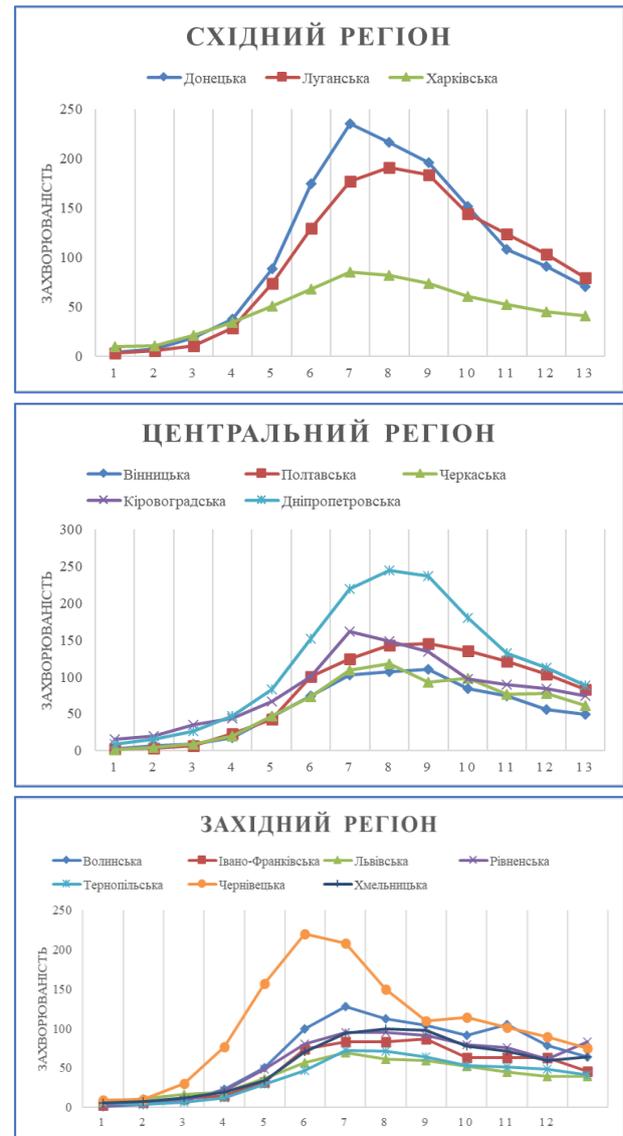


Рис. 1. Захворюваність по регіонах України

3) міра, що побудована для оцінювання близькості між характеристиками динаміки часових рядів. Опис динаміки часових рядів нормованих вимірювань проводився з застосуванням перших різниць з лагом 1: $\Delta x_{ir} = x_{i,r+1} - x_{i,r}$ ($i = \overline{1, m}, r = \overline{1, n-1}$). Нормалізація значень часових рядів дозволяє не звертати увагу на розкиданість спостережень у просторі, а досліджувати тільки поведінку рядів. Для побудови цієї міри введемо позначення: $L_{ir} = \begin{cases} 1, & \Delta x_{ir} \geq 0; \\ 0, & \Delta x_{ir} < 0. \end{cases}$

Очевидно, це перетворення часових рядів відображає динаміку процесу, а точніше, вказує

інтервали неспадання. В цих позначеннях міра $D_3(\Delta X_i, \Delta X_j)$ має такий вигляд:

$$D_3(\Delta X_i, \Delta X_j) = \frac{1}{n-1} \sum_{r=1}^{n-1} L1_{ir} \oplus L1_{jr},$$

де $L1_{ir} \oplus L1_{jr}$ – покомпонентне додавання за модулем 2.

Додавання за модулем 2 є логічною операцією, результат виконання якої буде істинним тоді і тільки тоді, коли лише один з аргументів істинний. Іншими словами, при обчисленні міри близькості $D_3(\Delta X_i, \Delta X_j)$ між двома рядами враховуються ситуації комбінацій поведінки часових рядів «спадає-зростає» або «зростає-спадає».

Дендрограма виконання процедури $\{IP, D_2(ST_i, ST_j)\}$ в якості характеристик об'єктів дослідження – часових рядів – обрано найпростіші статистичні характеристики, приведена на рис. 2.

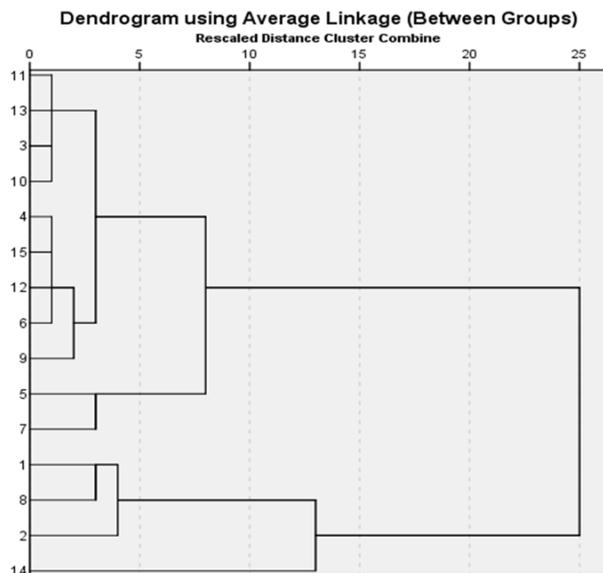


Рис. 2. Дендрограма процесу злиття процедури кластеризації $\{IP, D_2(ST_i, ST_j)\}$

В результаті проведених обчислень маємо два кластери, що складаються з 11 та 4 об'єктів. Для об'єднання всіх об'єктів в один кластер алгоритму знадобилось 25 кроків.

Схожі результати були отримані також при проведенні процедур кластеризації:

$$\{IP, D_1(X_i, X_j)\};$$

$$\{IP, D_3(\Delta X_i, \Delta X_j)\}.$$

Для часових рядів, що розглядались, аналіз дендрограм усіх трьох процедур кластеризації виявив таке:

1. Оптимальна кількість кластерів – 2.
2. Поділи областей регіонів України за кластерами суттєвих відмінностей не мають.
3. Для отримання найкращого поділу часових рядів на кластери в кожній процедурі використаний критерій мінімальності значення співвідношення F_0/F_1 .

Висновки

В роботі були проведені процедури кластерного аналізу коротких часових рядів з трьома мірами близькості:

Евклідова відстань;

міра близькості $D_2(ST_i, ST_j)$, що була побудована на статистичних характеристиках рядів;

міра близькості $D_3(\Delta X_i, \Delta X_j)$, що характеризує динаміку числових рядів.

Як свідчать результати виконання процедур, міри близькості $D_2(ST_i, ST_j)$ і $D_3(\Delta X_i, \Delta X_j)$ в застосуванні процедур кластеризації дають практично однаковий результат поділу множини $\Omega = \{X_1, X_2, \dots, X_m\}$ часових рядів на кластери. Це дозволяє застосовувати їх в задачах кластеризації коротких часових рядів і дає впевненість також в надійному застосуванні в якості міри близькості характеристик мінливості часових рядів як послідовності різниць другого порядку.

Результати кластеризації розглянутих числових рядів захворюваності обраних трьох регіонів України у подальшому будуть проаналізовані з ціллю виявлення причин саме такого об'єднання областей регіонів у кластери, а також причин аномальних значень захворюваності деякими шкірними хворобами в окремих містах регіонів. Як правило, ці причини пов'язані з відхиленнями екологічних показників повітря, води тощо. Як саме та яка величина ризиків захворюваності, треба буде встановлювати в окремих дослідженнях.

Конфлікт інтересів. Автори декларують, що не мають конфлікту інтересів стосовно даного дослідження, в тому числі фінансового, особистісного характеру, авторства чи іншого характеру, що міг би вплинути на дослідження та його результати, представлені в даній статті.

Використання засобів штучного інтелекту. Автори підтверджують, що не використовували технології штучного інтелекту при створенні представленої роботи.

СПИСОК ЛІТЕРАТУРИ

1. Kaufman L., Rousseeuw P.J. Finding groups in data: An introduction to cluster analysis. – Wiley, 2009. – 355 p. ISBN 0470317485, DOI: <https://doi.org/10.1002/978047031680>
2. Weiss S.M. A novel approximation to dynamic time warping allows anytime clustering of massive time series datasets. *Proceedings of the 2012 SIAM International Conference on Data Mining*. – pp. 999-1010, doi: <https://doi.org/10.1137/1.9781611972825.86>.
3. Giusti R., Batista G.E. An empirical comparison of dissimilarity measures for time series classification. *Proceedings of the Brazilian Conference on Intelligent Systems (BRACIS)*. – 2013. – pp. 82-88, doi: <https://doi.org/10.1109/BRACIS.2013.2>

4. Ding H. et al. Querying and mining of time series data: experimental comparison of representations and distance measures. *Proc. of the VLDB Endowment*. – 2015. – vol. 1, issue 2. – pp. 1542-1552, doi: <https://doi.org/10.14778/1454159.1454226>
5. Dieng S., Michel P., Guindo A., Sallah K., Ba E.H., Cissé B., Carrieri M.P., Sokhna C., Milligan P., Gaudart J. Application of Functional Data Analysis to Identify Patterns of Malaria Incidence, to Guide Targeted Control Strategies. *Int J Environ Res Public Health*. – 2020 Jun 11. – 17(11):4168, doi: <https://doi.org/10.3390/ijerph17114168>
6. Vasilios Zarikas, Stavros G. Pouloupoulos, Zoe Gareiou, Efthimios Zervas. Clustering analysis of countries using the COVID-19 cases dataset. *Data in Brief*. – Vol. 31, 2020, 105787, doi: <https://doi.org/10.1016/j.dib.2020.105787>
7. M. C. Lucic, H. Ghazzai, C. Lipizzi and Y. Massoud, "Integrating County-Level Socioeconomic Data for COVID-19 Forecasting in the United States", in *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 2, pp. 235-248, 2021, doi: 10.1109/OJEMB.2021.3096135, doi: <https://doi.org/10.1109/OJEMB.2021.3096135>.
8. Yeongryeol Choi, Nahyeon An, Seokyoung Hong, Hyungtae Cho, Jongkoo Lim, In-Su Han, Il Moon, Junghwan Kim, Time-series clustering approach for training data selection of a data-driven predictive model: Application to an industrial bio 2,3-butanediol distillation process, *Computers & Chemical Engineering*, Volume 161, 2022, 107758, ISSN 0098-1354, doi: <https://doi.org/10.1016/j.compchemeng.2022.107758>.
9. Кучанський, О. Ю., Білощицький, А. О. (2015). Selective pattern matching method for time-series forecasting. *Eastern-European Journal of Enterprise Technologies*, 6(4(78)), 13-18. URL: https://www.researchgate.net/publication/288323553_Selective_pattern_matching_method_for_time-series_forecasting.
10. Laurinec, Peter and Lucká, Mária. "Clustering-based forecasting method for individual consumers electricity load using time series representations", *Open Computer Science*, vol. 8, no. 1, 2018, pp. 38-50, doi: <https://doi.org/10.1515/comp-2018-0006>.
11. Perova I., Bodyanskiy Ye. Adaptive fuzzy clustering based on Manhattan metrics in medical and biological applications. *Вісник національного університету «Львівська політехніка»*, № 826, 2015. – С. 8-12. URL: https://sci.lubgd.edu.ua/bitstream/123456789/1053/1/Vis_826_Komp-nauky_min.pdf.
12. Song, C., Pei, T., Wang, X., Liu, Y., Ma, J., & Zhou, D. (2022). Dynamic characteristics of the COVID-19 epidemic in China's major cities. *Annals of GIS*, 28(4), 445-456, doi: <https://doi.org/10.1080/19475683.2022.2026468>.
13. T. Warren Liao, Clustering of time series data – a survey, *Pattern Recognition*, Volume 38, Issue 11, 2005, Pages 1857-1874, ISSN 0031-3203, doi: <https://doi.org/10.1016/j.patcog.2005.01.025>.
14. Солошенко Е. М., Волкославська В. М., Гутнев О. Л. Динаміка розповсюдженості та захворюваності на поширені дерматози в Україні і Харківському регіоні за останні 10 років. *Дерматологія та венерологія*. – 2014. – № 1. – С. 68-77. URL: http://nbuv.gov.ua/UJRN/dtv_2014_1_11.

Received (Надійшла) 14.11.2025

Accepted for publication (Прийнята до друку) 21.01.2026

Publication date (Дата публікації) 27.02.2026

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

Чікіна Наталія Олександрівна – кандидат технічних наук, доцент, професор кафедри вищої математики, Національний технічний університет «Харківський політехнічний інститут», Харків, Україна.

Chikina Nataliia – Candidate of Technical Sciences, Associate Professor, Professor of the Department of Higher Mathematics, National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine;

e-mail: nataliia.chikina@khp.edu.ua; ORCID Author ID: <http://orcid.org/0000-0002-0643-1771>.

Антонова Ірина Володимирівна – кандидат технічних наук, доцент, доцент кафедри вищої математики, Національний технічний університет «Харківський політехнічний інститут», Харків, Україна;

Antonova Iryna – Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Higher Mathematics, National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine;

e-mail: iryana.antonova@khp.edu.ua; ORCID Author ID: <http://orcid.org/0000-0002-1268-8223>.

Comparative analysis of time series

Nataliia Chikina, Iryna Antonova

Abstract. Relevance. Statistical data on most social, demographic, medical, and environmental indicators are typically collected on an annual basis. Such series are available only for small amounts of data over several years, i.e., they are short time series. Most existing clustering methods are not suitable for analysing short time series. Therefore, it is important to present information about numerical series in such a way that it is sufficient for the purposes of clustering short time series. Recently, publications have appeared in the scientific literature with recommendations to provide additional information contained in the statistical indicators of time series measurements, as well as indicators that describe the dynamics of time series, their variability, but do not depend on time. The task of selecting measures of proximity of series remains relevant. Existing studies on the comparative analysis of proximity measures tend to suggest that Euclidean distance ranks first in terms of quality in time series clustering tasks. **Object of study:** short time series of data on the prevalence of certain types of skin diseases in the regions of Ukraine. **Purpose of the article:** comparative analysis of the results of clustering short time series obtained using different proximity measures. **Research results.** The study examines time series of data on the number of cases of various skin diseases in the regions of Ukraine over a specified period. The results of clustering such time series make it possible to perform a comparative analysis of data and identify groups of regions of Ukraine with a «similar» situation in this regard. **Conclusions.** The results obtained can be used in further research to identify the causes of the corresponding levels of morbidity, such as water quality, air quality, and environmental pollution. To solve this problem, in addition to the Euclidean distance measure, the authors use a measure of proximity based on the main statistical characteristics of the time series and a measure of proximity for the dynamics indicators of the studied series, namely, the series of their first differences. A hierarchical procedure was selected from among the clustering methods for time series. A comparative analysis of the results obtained was carried out.

Keywords: mathematical modelling; short time series; comparative analysis; hierarchical cluster analysis procedure; measures of proximity; dendrogram; quality functional.