

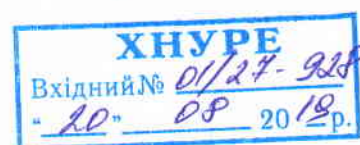
ВІДГУК
офіційного опонента
на дисертаційну роботу Кобиліна Іллі Олеговича
**«Нечітка кластеризація часових рядів в
інтелектуальному аналізі потоків даних»**,
що подана на здобуття наукового ступеня кандидата
технічних наук за спеціальністю 05.13.23 – системи та
засоби штучного інтелекту

Актуальність теми дисертаційної роботи. Сучасні методи обробки даних у вигляді часових рядів успішно застосовуються для вирішення багатьох прикладних задач, але, в той же час, використання подібних методів часто стає неефективним.

Наприклад, при обробці часових рядів в онлайн режимі часто буває невідома природа нових даних, що надходять на обробку, а системі необхідно адаптивно пристосовуватися до них в складних умовах невизначеності, нелінійності, стохастичності, а також наявності завад різного типу та багатовимірності.

Тому розробка онлайн методів нечіткої кластеризації нерівномірно квантованих асинхронних нестационарних часових рядів, чому присвячена дисертаційна робота Кобиліна І. О., є на сьогоднішній день актуальною. У роботі здійснено спробу створення нових методів кластеризації часових рядів, поєднуючи переваги використання класичних методів кластеризації у рамках методології інтелектуального аналізу даних.

Ступінь обґрунтованості наукових положень, висновків та рекомендацій. У дисертаційній роботі досить коректно здійснено постановку задачі нечіткої кластеризації часових рядів, що призначена для інтелектуального аналізу даних у послідовному онлайн режимі надходження потоку даних. Основні припущення, що покладені в основу проведених теоретичних досліджень, є логічно викладеними, обґрунтованими і коректними.



Результативність запропонованих автором методів кластеризації часових рядів переконлива і у достатній мірі обґрунтована за допомогою аналітичних досліджень, імітаційного моделювання, а також експериментів на реальних даних.

Достовірність отриманих результатів забезпечується коректним використанням теорії штучних нейронних мереж, апарату нечіткої логіки, теорії оптимізації, імітаційного моделювання, а також результатами використання розроблених методів їх навчання для розв'язання прикладних задач інтелектуального аналізу даних.

Новизна результатів дисертації. Загалом можна погодитися, що отримані дисертантом основні результати дослідження містять наступні конкретні ознаки наукової новизни:

1. Вперше запропоновано метод кластеризації, який неохильний до ефекту концентрації норм, що дозволяє вирішувати задачу кластеризації в онлайн режимі за умов перетину класів та асинхронних нерівномірно квантованих часових рядів за рахунок використання спеціальної цільової функції нечіткої кластеризації.

2. Вперше запропоновано послідовний онлайн метод кластеризації багатовимірних часових рядів, що базується на апараті гібридних систем обчислювального інтелекту, який дозволив вирішувати задачу кластеризації даних, які послідовно надходять на обробку з нерівномірними тактами квантування.

3. Отримав подальший розвиток метод адаптивної кластеризації, що базується на методах ймовірнісної та можливісної кластеризації коротких часових рядів, які, у свою чергу, засновані на метриці спеціального вигляду, що дозволяє значно спростити чисельну реалізацію методу, за рахунок використання метрики на основі тангенсів кутів нахилу, що на відміну від

відомих методів вирішує задачу кластеризації нерівномірно квантованих часових рядів.

4. Отримав подальший розвиток метод робастної адаптивної ідентифікації нестационарних часових рядів в онлайн режимі надходження потоку даних, який характеризується простотою обчислювальної реалізації та вирішує задачу обробки даних, що збурені аномальними викидами, за рахунок використання введеної модифікації критерія Гемана-МакКлюра.

Практичне значення результатів дисертаційної роботи. Методи кластеризації часових рядів, запропоновані в дисертаційній роботі, можуть бути використані для обробки даних, поданих у вигляді таблиць «об'єкт-властивість» або часових послідовностей. Проведені експериментальні дослідження довели, що запропоновані методи можуть бути успішно використані для розв'язання прикладних задач, зокрема, метод нечіткої кластеризації нерівномірно квантованих асинхронних часових рядів та його застосування для вирішення задач обробки медичних даних у моніторингових системах.

Рекомендується продовжити дослідження з розроблення та використання методів, моделей та інструментальних засобів інтелектуального аналізу даних у формі часових рядів у Харківському національному університеті радіоелектроніки, споріднених ВНЗ та у прикладних системах аналізу даних.

Зв'язок з науковими програмами, планами, темами. Дисертаційна робота виконана в рамках держбюджетних НДР: «Нейро-фаззі системи для поточної кластеризації та класифікації послідовностей даних в умовах їх спотворення відсутніми і аномальними спостереженнями» (№ДР 0113U000361); «Динамічний інтелектуальний аналіз послідовностей нечіткої інформації в умовах суттєвої невизначеності на основі гібридних систем обчислювального інтелекту» (№ДР 0116U002539), які виконувалися

у Харківському національному університеті радіоелектроніки, згідно з наказами Міністерства освіти і науки України за результатами конкурсного відбору проєктів наукових досліджень, у яких автор брав участь як виконавець.

Повнота викладення основних результатів дисертації в наукових виданнях. Основні положення дисертаційної роботи опубліковані у 12 наукових роботах: 1 розділ у колективній монографії, що входить до наукометричної бази Scopus; 4 статті у виданнях, які зазначені в переліках фахових видань України з технічних наук та 1 стаття за кордоном, що входить до наукометричної бази Scopus; 7 публікацій у матеріалах конференцій (2 включено до наукометричної бази даних Scopus).

Відповідність дисертації та автореферату встановленим вимогам. За результатами загального аналізу дисертаційної роботи можна стверджувати, що вона є завершеним і цілісним дослідженням з логічним викладенням матеріалу та чіткою структурою. Матеріал викладено з використанням сучасної науково-технічної термінології.

Згідно з вимогами пп. 9, 11-14 «Порядку присудження наукових ступенів» (Постанова КМУ №567 від 24 липня 2013 р.) стиль викладення матеріалів досліджень, наукових положень та результатів імітаційного моделювання забезпечує доступність сприйняття дисертаційної роботи.

Зміст автореферату повністю розкриває сутність основних наукових положень дисертаційної роботи. Автореферат лаконічно та адекватно відображає отримані результати. Оформлення автореферату за обсягом, структурою та змістом відповідає чинним вимогам.

Зауваження по дисертаційній роботі. Дисертаційна робота не вільна від недоліків, до яких можна віднести нижченаведені.

1. Як виходить з тексту дисертаційної роботи та формулювань наукової новизни, особливістю запропонованих автором методів є їх

можливе застосування для кластеризації нерівномірно квантованих числових рядів. В той же час, у першому розділі не приділено уваги проблемам та особливостям, що викликані саме нерівномірним квантуванням чи асинхронністю часових рядів.

2. При обробці нерівномірно квантованих часових рядів типовим є використання інтерполяції для трансформації їх у рівномірно квантовані часові ряди та застосування до них відомих методів. У роботі не розглянуто цю можливість, що не дає змоги зрозуміти, чи є методи, запропоновані автором, кращими за подібний підхід.

3. У роботі недостатньо наведено практичне значення застосування саме нечіткої кластеризації, роль нечіткості є незрозумілою. Наприклад, імітаційне моделювання методів адаптивної можливісної нечіткої кластеризації коротких часових рядів у підрозділі 5.2 проведено зі значенням фаззифікатора, що дорівнює $\beta=1.1$, тобто фаззифікація майже відсутня.

4. Запропоновані у роботі методи характеризуються обранням значної кількості параметрів, таких як швидкість навчання, параметр фаззифікації, кількість кластерів, кількість моделей у ансамблі тощо. Хоча б якесь обґрунтування значень вказаних параметрів у практичній частині також відсутнє.

5. Із матеріалів дисертаційного дослідження недостатньо точно зрозумілими є області та особливості використання кожного з запропонованих методів окремо. Наприклад, вказано, що метод є стійким до ефекту концентрації норм, але з практичної точки зору цю властивість не доведено, незрозуміло, чи зможуть представлені методи обробляти вибірки з різними довжинами часових рядів.

6. П'ятий розділ, що містить результати імітаційного моделювання, не відображає у повній мірі практичних переваг запропонованих у

дослідження методів. Замість застосування гібридної адаптивної моделі ідентифікації зашумленого сигналу з незрозумілих причин використано ансамбль таких моделей. Більшість рисунків (наприклад, рис. 5.3 – 5.5, рис. 5.9 – 5.11) потребують додаткових текстових пояснень, хоча б підписів, наведених у легенді. Для однієї з сукупностей методів наведено для порівняння одразу 3 індекси оцінювання якості кластеризації: індекс силуета, Цалінські-Харабаша та Девіса-Боулдіна, хоча цілком достатньо було б хоча б одного з них, використання всіх не є необхідним. Більше того, цей розділ перенасичений описом медичних параметрів, які є неважливими для розуміння роботи запропонованих методів.

7. Робота містить недоліки оформлення, наприклад:

- загалом матеріали перевантажено великою кількістю математичних формул, що не полегшує розуміння запропонованих ідей та методів;

- формула, наведена на с. 47 дисертації є ідентичною до формули (2.12) якщо $\sigma^2 = 1$ (окрім знаку множення, який чомусь відрізняється), однак відповідні вигляди функцій, наведені на рис. 2.2 та 2.3 не співпадають;

- підписи під рисунками 2.3 – 2.5 не відповідають посиланням на них у тексті роботи;

- посилання на індекс Цалінські-Харабаша [108] є некоректним.

Вважаю, що наведені зауваження суттєво не впливають на позитивну оцінку проведених досліджень.

Загальні висновки по дисертаційній роботі. Тема і зміст дисертаційної роботи Кобиліна І. О. відповідають наступним пунктам паспорта спеціальності 05.13.23 – системи та засоби штучного інтелекту:

п. 11 «Створення математичних моделей на принципах нечіткої логіки для застосування в системах штучного інтелекту»;

п. 14 «Аналіз, синтез та моделювання нейронних мереж, розроблення методів їх проектування, оптимізації та навчання».

Дисертаційна робота є завершеним науковим дослідженням, що містить науково обґрунтовані результати, які мають важливе прикладне значення. Автореферат оформлений згідно з вимогами, він повністю описує суть отриманих результатів та відповідає змісту дисертації.

Дисертація відповідає вимогам п. 9, 11, 12, 14 «Порядку присудження наукових ступенів» (Постанова КМУ № 567, від 24 липня 2013 р.) щодо кандидатських дисертацій, а її автор, Кобилін Ілля Олегович, заслуговує на присудження йому наукового ступеня кандидата технічних наук за спеціальністю 05.13.23 – системи та засоби штучного інтелекту.

Офіційний опонент:

Доцент
кафедри інформатики та комп'ютерної техніки
Харківського національного економічного
університету імені Семена Кузнеця,
кандидат технічних наук, доцент


О. В. Гороховатський

