

## ВІДГУК

### офіційного опонента

на дисертаційну роботу **Кобиліна Іллі Олеговича**

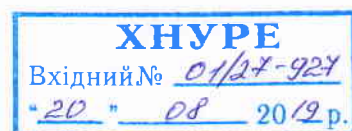
**«Нечітка кластеризація часових рядів**

**в інтелектуальному аналізі потоків даних»**,

подану до захисту на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.23 – системи та засоби штучного інтелекту

**Актуальність теми дисертаційної роботи.** Задача обробки даних, що надходять до обчислювальної системи послідовно в оперативному режимі є актуальною, оскільки останні роки характеризуються бурхливим розвитком таких напрямів штучного інтелекту, як Data Stream Mining та Dynamic Data Mining, що підтверджується зростаючою кількістю публікацій та міжнародних конференцій у цій сфері. Особливістю оброблюваних даних у цих задачах є їхня розподіленість у часі, а характерною особливістю задач, що вирішуються при обробці часових рядів, є їхня багатовимірність, те, що обсяг вибірки не є фіксованим і зростає з плином часу, а характер спостережень може змінюватись. Ключову роль в обробці потоків даних відіграє кластерний аналіз. Варто відзначити, що на цей час існує багато різних методів кластеризації, але вони або працюють з вибіркою фіксованого обсягу, або є занадто громіздкими з обчислювальної точки зору, або можуть виконувати тільки чітку кластеризацію. Тому на сьогодні важливою і актуальною є розробка методів нечіткої кластеризації часових рядів в оперативному режимі, де потрібна ефективність та швидкодія при обробці даних для вирішення цілої низки задач інтелектуального аналізу даних. Саме цій темі присвячена дисертаційна робота І. О. Кобиліна.

**Ступінь обґрунтованості наукових положень, висновків та рекомендацій.** Автором проведено огляд та аналіз відомих методів обробки



часових рядів, що досить широко використовуються для розв'язання задач інтелектуального аналізу даних. Крім того, виявлено недоліки таких систем та зроблено огляд методів створення систем обчислювального інтелекту, які покращують якість обробки даних, що надходять у послідовному оперативному режимі. Ці огляд та аналіз дозволили авторів сформулювати мету роботи, науково-практичне завдання, виділити об'єкт та предмет дослідження.

Основні припущення, що покладені в основу теоретичних досліджень, є коректними, а імітаційне моделювання розроблених методів нечіткої кластеризації часових рядів систем підтвердило їхню ефективність у розв'язанні задач інтелектуального аналізу даних.

**Достовірність отриманих результатів.** Достовірність викладених у дисертаційній роботі наукових положень, результатів і висновків, зроблених здобувачем, підтверджується даними, що були отримані при імітаційному моделюванні, впровадженням результатів роботи, а також апробацією на міжнародних наукових конференціях.

**Наукова новизна результатів дисертації.** Аналіз дисертаційної роботи дозволяє зробити висновок, що автором у процесі досліджень отримані такі наукові результати.

1. Вперше запропоновано метод кластеризації, який несхильний до ефекту концентрації норм, що дозволяє вирішувати задачу кластеризації в online 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.

**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 проектів наукових досліджень.

**Практичне значення результатів дисертаційної роботи.** Отримані теоретичні результати були досліджені експериментально на тестових і реальних даних, де довели свою перевагу над відомими методами. Запропоновані в роботі методи передобробки та обробки часових рядів можуть бути використані в різних областях, де дані представлені в числовій формі у вигляді таблиць «об'єкт-властивість» або у вигляді часових послідовностей в online режимі. Запропоновані архітектури і методи їх

навчання довели свою ефективність при розв'язанні таких практичних задач:

- моніторинг медичних даних в online режимі у ТОВ «Інфобуд», м. Харків;
- моніторинг медичних даних та їх кластеризація за допомогою запропонованого методу нечіткої кластеризації у ТОВ «Сайтосс», м. Харків.

Усі впровадження підтверджено відповідними актами.

**Повнота викладу основних результатів у наукових виданнях та апробація.** За результатами досліджень здобувачем опубліковано загалом 12 наукових робіт: з них 1 розділ у колективній монографії, що входить до наукометричної бази SCOPUS; 1 стаття за кордоном, що входить до наукометричної бази SCOPUS; 3 статті у періодичних фахових виданнях з технічних наук, що включені до переліків МОН України, 7 публікацій у працях наукових конференцій, 2 з яких включено до наукометричної бази даних SCOPUS. Усі конференції присвячені питанням обчислювального інтелекту та інтелектуального аналізу даних.

**Відповідність змісту автореферату основним положенням дисертації.** Оформлення автореферату за своїм обсягом, структурою та змістом відповідає чинним вимогам. Зміст автореферату ідентичний змісту основних положень дисертації, автореферат адекватно відображає результати дисертації.

**Відповідність дисертації встановленим вимогам.** Дисертаційна робота є завершеним і цілісним дослідженням, її матеріал є досить добре структурованим і логічно викладеним. Роботу написано коректно з використанням сучасної науково-технічної термінології.

Оформлення дисертації відповідає встановленим вимогам до кандидатських дисертацій згідно з пп. 9, 11–14 «Порядку присудження наукових ступенів» (Постанова КМУ №567 від 24 липня 2013 р.), а також вимогам МОН України до дисертацій на здобуття наукового ступеня кандидата технічних наук. Стиль викладення матеріалів досліджень, наукових

положень, результатів роботи систем при імітаційному моделюванні забезпечує їх адекватне і належне сприйняття.

**Зауваження по дисертаційній роботі.** Серед недоліків дисертації слід зазначити такі.

1. У першому розділі роботи недостатню увагу приділено аналізу існуючих методів кластеризації часових рядів та їхніх недоліків.

2. У другому розділі варто було навести аналіз та детальний опис адаптивного методу фільтрації нестационарних часових рядів.

3. Варто було б детальніше розглянути недоліки використання ймовірнісного підходу до нечіткої кластеризації.

4. При використанні WTM та WTA правил Т. Кохонена надано замало інформації щодо їх використання у методах кластеризації.

5. Не зрозуміло, як саме обирається параметр фазифікації для обробки часового ряду в online режимі.

6. Варто було детальніше розглянути результати ідентифікації часових рядів при їх зашумленості.

7. Для оцінки якості кластеризації під час імітаційного моделювання варто було використати й інші індекси, окрім індексів силуета та Девіса-Болдуїна.

Загалом, вказані недоліки не впливають на високу оцінку виконаних автором досліджень.

**Загальні висновки по дисертаційній роботі.** Дисертаційна робота Кобиліна І. О. є завершеним науковим дослідженням, у якій розв'язано актуальну на цей час науково-практичну задачу, а саме розробку методів обробки часових рядів, які дозволяють обробку в online режимі. Виконані в дисертаційній роботі дослідження та отримані наукові результати відповідають паспорту 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». Автореферат оформлений згідно з чинними вимогами, що висуваються до кандидатських дисертацій, він повністю відповідає змісту дисертації і описує суть одержаних результатів та висновків у дисертаційній роботі.

Дисертація відповідає вимогам пп. 9, 11, 12, 13, 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:**

Завідувач кафедри програмних засобів

Національного університету "Запорізька політехніка",

доктор технічних наук, професор

С. О. Субботін

**Підпис проф. С.О. Субботіна засвідчую.**

Вчений секретар Національного університету

"Запорізька політехніка",

кандидат соціологічних наук, доцент



В. В. Кузьмін