

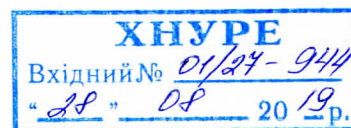
## **ВІДЗИВ офіційного опонента**

доцента кафедри інформатики та комп'ютерної техніки Харківського національного економічного університету імені Семена Кузнеця, кандидата технічних наук, доцента Гороховатського Олексія Володимировича на дисертаційну роботу Жернової Поліни Євгеніївни «Нечітка кластеризація потоків даних за умов невідомої кількості кластерів», що подана на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.23 – системи та засоби штучного інтелекту

**Актуальність теми дисертаційної роботи.** Задачі інтелектуального аналізу даних, до яких відноситься задача кластеризації, набули значного поширення останніми роками завдяки їх впровадженню у повсякденне життя. Більшість сучасних методів кластеризації даних розроблено у припущенні, що межі кластерів є чіткими (crisp partition), що обмежує їх застосування для нечітких даних, які часто сприймаються людьми більш природньо. Окрім цього, традиційно вважається необхідним навчити метод кластеризації на якихось даних, після чого застосовувати його на практиці (пакетний режим). Але багато різних даних є достатньо динамічними для накопичення, тому що до моменту застосування навченої на них моделі вони можуть застаріти. Більш логічним у даному випадку є застосування методів онлайн обробки потоку даних, які наділені властивістю динамічно змінюватися і навчатися під час надходження вимірюванні за вимірюванням.

Таким чином, на сьогоднішній день актуальною є наукова задача розробки нових методів нечіткої ядерної кластеризації, призначених для обробки даних в on-line режимі, коли дані надходять на обробку послідовно, одне за одним, а кластери можуть перетинатися і мати довільну форму.

**Ступінь обґрунтованості наукових положень, висновків та рекомендацій.** Запропоновані в роботі методи нечіткої кластеризації на основі ансамблю ядерних функцій та нейро-фаззі системи обчислювального інтелекту



призначені для online кластеризації потоку інформації в умовах невизначеності чи відсутності інформації як про кількість кластерів, так і про їх форму. Введені нейро-фаззі системи досить прості в чисельній реалізації і дозволяють вирішувати задачі інтелектуального аналізу потоків даних. Ці методи дозволили встановити закономірності формування відповідної реакції організму на об'єднаний вплив екологічних чинників; використання запропонованих підходів допомогло визначити значення біологічних ефектів сполученої дії електромагнітного випромінювання та позитивних низьких температур при аналізі результатів у рамках науково-дослідної роботи, що фінансувалася Міністерством охорони здоров'я України, «Встановити механізми адаптації до сполученої дії хімічних та фізичних чинників навколишнього середовища».

Також основні результати дисертаційної роботи використовуються у навчальному процесі Харківського національного університету радіоелектроніки на кафедрі системотехніки в курсі «Нейросистеми та генетичні алгоритми».

**Зв'язок роботи з науковими програмами, планами, темами.** Дисертаційна робота виконана в рамках держбюджетних НДР: «Динамічний інтелектуальний аналіз послідовностей нечіткої інформації за умов суттєвої невизначеності на основі гібридних систем обчислювального інтелекту» (№ДР 0116U002539); «Глибинні гібридні системи обчислювального інтелекту для аналізу потоків даних та їх швидке навчання» (№ДР 0119U001403). У рамках зазначених НДР здобувачем розроблені методи синтезу ансамблів нечіткої кластеризації, які призначені для обробки даних в онлайн режимі, коли дані надходять на обробку послідовно, а кластери можуть перетинатися у просторі ознак та мати довільну форму.

**Достовірність отриманих результатів** забезпечується коректним використанням теорії штучних нейронних мереж, теорії інтелектуального

аналізу даних, імітаційного моделювання та результатами експериментів на реальних об'єктах.

### **Новизна результатів дисертації.**

Основні наукові результати полягають в розробці методів кластеризації потоків даних на основі ансамблевого підходу з використанням самоорганізовних карт Т. Кохонена та методу С-середніх для вирішення проблеми ефективного динамічного інтелектуального аналізу потоків даних.

Отримана новизна:

1. Вперше запропоновано ансамбль самоорганізовних карт Т. Кохонена, який характеризується використанням онлайн методу К-середніх, що дозволяє кластеризувати дані за умов апріорі невідомої кількості класів.

2. Вперше запропоновано ансамбль нейро-фаззі самоорганізовних карт Т. Кохонена, який характеризується використанням модифікованого онлайн методу нечітких С-середніх, коли апріорі невідома кількість та форма кластерів, що дозволяє кластеризувати потоки даних за умов лінійної нероздільності класів, які довільним чином перетинаються у просторі ознак.

3. Удосконалено ансамбль ядерних самоорганізовних карт Т. Кохонена, який відрізняється від аналогів введенням додаткового ядерного шару для підвищення розмірності вхідного простору, що дозволяє кластеризувати потоки даних за умов, коли кластери є лінійно нероздільними.

4. Удосконалено ансамбль самоорганізовних нечітких карт Т. Кохонена, який відрізняється від аналогів одночасним використанням процедури імовірнісної та можливісної кластеризації потоків даних, що дозволяє підвищити рівень якості кластеризації потоків даних.

**Повнота викладення основних результатів дисертації в наукових виданнях.** Основні положення дисертаційної роботи опубліковані у 14 наукових роботах: 6 статей у фахових виданнях України з технічних наук; частина монографії, що індексується у наукометричній базі Scopus; 8 тез

доповідей на наукових конференціях (2 з яких індексуються у наукометричній базі Scopus).

Низка публікацій входить до міжнародних наукометричних та реферативних наукових баз (Index Copernicus, INSPEC, BASE та інші).

### **Апробація наукових результатів**

Результати пройшли апробацію на низці конференцій як в Україні, з яких можна виділити Міжнародну школу-семінар «Теорія прийняття рішень»; Міжнародну наукову конференцію «Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту»; IEEE Second International Conference on Data Stream Mining & Processing; Міжнародна науково-практична конференція «Інформаційні технології та системи.

### **Рекомендації щодо використання результатів дисертації.**

Запропоновані автором результати доцільно використати для підвищення якості кластеризації потоків даних за умов невизначеної кількості кластерів та їх форми у різних сферах діяльності, таких як, медицина, техніка та інші. Отримані результати також можуть бути використані у навчальному процесі при підготовці спеціалістів у галузі інтелектуальної обробки даних.

### **Відповідність дисертації встановленим вимогам**

При загальній оцінці кандидатської дисертаційної роботи слід зазначити, що вона є завершеним і цілісним дослідженням з чіткою структурою і логічним викладом матеріалу, узагальнює дослідження автора, написана сучасною науково-технічною мовою.

Оформлення дисертації проведено згідно з вимогами «Порядку присудження наукових ступенів і присвоєння вченого звання старшого наукового співробітника». Стиль викладення матеріальних досліджень, наукових положень, висновків і рекомендацій забезпечує доступність її сприйняття.

## **Відповідність змісту автореферату основним положенням дисертації**

Оформлення автореферату за своїм обсягом, структурою та змістом відповідає чинним вимогам «Порядку присудження наукових ступенів і присвоєння вченого звання старшого наукового співробітника». Зміст автореферату повністю розкриває зміст основних наукових положень дисертаційної роботи.

**Зауваження по роботі.** Матеріали, наведені у дисертаційній роботі та авторефераті, мають такі недоліки:

1. Матеріали досліджень не містять обґрунтування використаної міри подібності для навчання самоорганізованих мап та порівняння відстаней, що може бути дуже важливим у задачах кластеризації.

2. У розділі 2 запропоновано ансамбль самоорганізованих карт Т. Кохонена, який використовує онлайн метод К-середніх та є за суттю адаптивною модифікацією методу Х-середніх. Але порівняння ефективності цього методу виконано не з методом Х-середніх, а з методом К-середніх, який, на відміну від запропонованого метода, працює із відомою кількістю кластерів. Для дослідження запропонованого ансамблю використано набір даних "ірисі Фішера", однак це дослідження показує не ефективність власне ансамблю (яка залишилась незрозумілою, оскільки чисельні показники точності кластеризації не наведено), а тільки ефективність обрання кількості кластерів.

3. У розділі 6 наведено дослідження ефективності запропонованих ансамблів для різних наборів даних. Показано, що методи здатні визначати кількість кластерів коректно. Незрозумілою є якість кластеризації у порівнянні із ground truth чи такою, що була отримана із використанням state-of-the-art методів, оскільки обрані набори даних є надзвичайно популярними.

4. Оцінка якості кластеризації під час роботи із різними наборами даних проводиться із використанням різноманітних індексів, вибір яких для конкретної задачі не обґрунтований.

5. Відомі інші методи кластеризації за умов невідомої кількості кластерів, порівняння із ними у дослідженнях відсутнє. Існують також достатньо прості методи "gap statistics" та "метод ліктя" ("elbow method") для визначення оптимальної кількості кластерів, останній з яких є за суттю близьким до роботи запропонованих ансамблів. Вважаю доцільним порівняння запропонованих методів хоча б з якимись відомими.

6. У роботі недостатньо продемонстровано практичне значення нечіткості для обробки використаних наборів даних.

7. Робота містить технічні недоліки, наприклад:

– на рис. 1.4 наведено, вочевидь, не чотири кластери, а один і той самий зі зміщенням;

– значення індексу Ксі-Бені для методу К-середніх у таблиці 4.1 виглядають як помилки обчислення, тому що нескладно знайти підтвердження, що метод К-means досить непогано працює для зазначеного набору даних.

Але, незважаючи на вказані зауваження, здійснені в дисертаційній роботі дослідження є новими, змістовними і корисними.

**Загальні висновки по дисертаційній роботі.** Тема і зміст дисертації П.Є. Жернової відповідають паспорту спеціальності 05.13.23 – системи та засоби штучного інтелекту, зокрема п. 11 та п. 14. Дисертаційна робота є завершеним науковим дослідженням і містить нові науково обґрунтовані результати, що є суттєвим внеском у галузі динамічного інтелектуального аналізу медичних даних. Отримані наукові результати мають важливе значення для розв'язання практичних задач.

Автореферат оформлений згідно діючих вимог, що висуваються до кандидатських дисертацій, повністю розкриває сутність дисертації та коректно описує одержані наукові результати та висновки у дисертаційній роботі.

Дисертація відповідає вимогам п. 9, 11, 12, 14 «Порядку присудження наукових ступенів і присвоєння вченого звання старшого наукового співробітника» (Постанова КМУ № 567 від 24 липня 2013 р.) щодо кандидатських дисертацій, а її автор – Жернова Поліна Євгеніївна, заслуговує на присудження 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

