

## ВІДГУК

офіційного опонента на дисертаційну роботу **Повхана Ігоря Федоровича** “Методи та принципи побудови дерев класифікації дискретних об’єктів для інтелектуального аналізу даних”, подану на здобуття наукового ступеня доктора технічних наук за спеціальністю 05.13.23 – системи та засоби штучного інтелекту

**1. Актуальність теми дисертації.** Задачі, які об’єднуються тематикою розпізнавання дискретних об’єктів виникають у сучасному світі в усіх сферах економіки та соціального контенту діяльності людини, що приводить до необхідності побудови та дослідження математичних моделей інтелектуального аналізу даних, які є складовою теорії штучного інтелекту. На даний час не існує універсального підходу до розв’язання задач класифікації. Існують загальні методи класифікації, які забезпечують вирішення багатьох типів задач, їх використання характеризується великою чутливістю до специфіки задачі та предметної області застосування.

Методи дерев класифікації позбавлені значної частини недоліків класичних підходів розпізнавання та забезпечують ефективну роботу в задачах з даними довільних шкал, де інформація задається в природній формі. Тому актуальним є розроблення систем класифікації на основі дерев рішень, які володіють рядом корисних властивостей – зменшенням складності моделей, підвищенням точності класифікації та розширенням галузей застосування дерев рішень в інтелектуальному аналізі даних.

Виходячи з цього, актуальною науково-прикладною проблемою, якій присвячено дисертаційну роботу І.Ф. Повхана, є розвиток теорії аналізу та синтезу дерев рішень, розроблення моделей, методів, прикладного інструментарію інтелектуального аналізу даних на основі логічних та алгоритмічних дерев класифікації з більшою точністю, зменшеною складністю моделей та підвищеною ефективністю класифікації дискретних об’єктів.

**2. Ступінь обґрунтованості наукових положень, висновків і рекомендацій.** Викладені в дисертації основні наукові положення, висновки і результати, що отримані здобувачем, забезпечуються правильністю постановки задачі та їх подальшим теоретичним викладенням, які співпадають із одержаними практичними результатами, апробацією на конференціях і семінарах. Усі наукові положення ґрунтуються на детальному аналізі об’єкта та предмета дослідження. Детальне ознайомлення з роботою та представленими в ній положеннями, дозволяє виділити наукову новизну та сформулювати чіткий порядок викладення матеріалу.

Дисертація є завершеною науково-дослідною роботою. Її структура логічна, вона оформлена акуратно та згідно встановлених вимог. Автореферат повністю відображає зміст дисертації.

Дисертаційна робота включає вступ, 7 розділів, висновки, 3 додатки та список використаних джерел. Загальний обсяг дисертації складає 382 сторінки, 67 рисунків та 58 таблиць, список літератури з 312 найменуваннями.

**3. Наукова новизна одержаних результатів.** У дисертаційній роботі вирішено актуальну науково-прикладну проблему розвитку теорії аналізу та синтезу дерев рішень, розроблення моделей, методів, прикладного інструментарію інтелектуального аналізу даних на основі логічних та алгоритмічних дерев класифікації з більшою точністю, зменшеною складністю моделей та підвищеною ефективністю класифікації дискретних об'єктів.

При цьому отримано такі результати:

**вперше розроблено:**

- комплексний метод побудови деревоподібних моделей класифікації (структур АДК);
- метод T-опорних множин у задачах представлення дискретних об'єктів;
- метод побудови структур алгоритмічних дерев класифікації (АДК), які відрізняються модульним принципом побудови моделей;
- метод побудови обмежених за складністю структур алгоритмічних дерев класифікації;
- метод знаходження подібності конструкцій логічних дерев в задачах мінімізації їх структур;
- метод оцінки впливу процедури обрізки логічного дерева класифікації (ЛДК) за рахунок перестановки ярусів, рівнів у конструкції регулярного логічного дерева;

**отримали подальший розвиток:**

- методи структур логічних дерев класифікації, випадкових дерев класифікації, які забезпечують побудову якісно кращих структур дерев класифікації;
- метод оцінки збіжності процедури побудови моделей дерев класифікації для умов слабого та сильного розділення класів навчальної вибірки;

**удосконалено:**

- схеми дерев моделей класифікації.

**4. Зв'язок роботи з науковими програмами, планами, темами.** Тема дисертаційної роботи відповідає науковим напрямкам кафедр факультету інформаційних технологій ДВНЗ «Ужгородський національний університет». Основу дисертаційної роботи складають результати теоретичних і практичних досліджень, які були використані у таких науково-дослідних програмах, темах, проєктах:

– кафедри інформатики та фізико-математичних дисциплін ДВНЗ «Ужгородський національний університет»: «Обробка великих масивів інформації за допомогою логіко-математичних методів» (номер державної реєстрації 0119U100733);

– кафедри програмного забезпечення систем ДВНЗ «Ужгородський національний університет»: «Методи та засоби програмної інженерії реалізації процесів аналітики великих масивів даних на базі інформаційно-технічних платформ» (номер державної реєстрації 0119U100703);

– науково-дослідної роботи «Моделювання та передбачення надзвичайних ситуацій в Карпатському регіоні та країнах Центрально-Східної Європи», номер державної реєстрації роботи – 0106V00285, категорія роботи – фундаментальні дослідження (КПКВ 2201020), 01 Фундаментальні дослідження з найважливіших проблем природничих, суспільних і гуманітарних наук;

– «Інноваційні методи навчання на підтримку партнерських відносин – InovEduc (2015–2017)» – грантового проекту № SVC01008 Норвезького державного фонду із солідарним бюджетом Словацької Республіки в рамках програми SK08 транскордонне співробітництво.

**5. Практична значущість одержаних результатів** полягає в тому, що запропоновані методи, моделі, алгоритми та програмні інструменти дають змогу:

– розробити схеми та програмні засоби для аналізу та синтезу моделей алгоритмічних дерев класифікації дискретних об'єктів;

– розробити схеми та програмні засоби для аналізу та декомпозиції побудованих моделей логічних дерев класифікації з можливістю виділення наборів класифікаторів;

– розробити програмний інструментарій побудови обмежених моделей дерев алгоритмів, що дає змогу зменшити час синтезу моделі АДК на 15-20%;

– розробити алгоритми та програмні засоби обчислення інформативності дискретних ознак у схемі синтезу моделі ЛДК;

– зменшити структурну складність моделей АДК на основі обмеженого методу синтезу дерев класифікації;

– отримати зменшення параметру структурної складності логічного дерева в процедурі обрізки ЛДК на основі схеми перестановки ярусів в 2,5–3 рази.

Низка побудованих моделей і схем АДК у межах даного дослідження використано ТОВ «ІНФОСФЕРА» та ТОВ «Медіа-Сервіс» (м. Ужгород), Управління економічного розвитку міста Ужгородської міської ради, Закарпатською обласною громадською організацією «Патріотичний оборонний – спортивний центр Вітязь». Додатково окремі положення дисертаційного дослідження використовуються в різних сферах навчального процесу факультету інформаційних технологій ДВНЗ «Ужгородський національний університет» при викладанні дисциплін: «Методи та засоби штучного інтелекту», «Програмування в системах абстрактних об'єктів і задачі штучного інтелекту», «Теорія алгоритмів», «Технології програмування та створення програмних продуктів», «Пакети прикладних програм», «Організація баз даних і знань».

## **6. Оцінка основного змісту дисертації.**

У вступі дисертантом обґрунтовано актуальність наукової проблеми та теми дисертаційної роботи, сформульовано мету і задачі, наведено використані методи дослідження, сформульовано наукову новизну та практичну цінність отриманих результатів і показано зв'язок проблеми з науковими програмами, планами та темами. Наведено дані про впровадження результатів роботи, їх апробацію, публікації та особистий внесок здобувача.

**У першому розділі** проведено аналіз деревоподібних моделей класифікації, наведено загальну постановку задачі навчання, виділено основні питання, що виникають на етапі навчання систем розпізнавання. Проаналізовано сучасний стан досліджень методів дерев рішень, висвітлено проблематику даної концепції, визначено шляхи подальших вдосконалень, перспективних досліджень моделей та методів дерев класифікації. Подано структурно-логічну схему досліджень.

**У другому розділі** наведено загальну концепцію дерев класифікації, яка забезпечує покриття масиву даних навчальної інформації за рахунок фіксації об'єктів вибірки в своїй структурі. Зафіксовано відсутність актуальних методів та алгоритмів, які би дозволили будувати ефективні конструкції логічних дерев на основі масивів даних великої розмірності, розглянуто окремих клас дерев класифікації – випадкові ЛДК. Розроблено алгоритмічну схему виправлення помилок ЛДК шляхом корекції структури дерева класифікації. Дано числову оцінку для фіксованого шляху структури довільного ЛДК максимальної кількості помилок класифікації всіх типів у процедурі розпізнавання. Досліджено питання можливості побудови ЛДК мінімальної складності на основі мінімального тесту та структурної складності максимального дерева класифікації.

**У третьому розділі** розроблено метод Т-опорних множин, який полягає у відборі, фіксації певного набору ознак разом зі своїми значеннями на основі інформації деякої початкової НВ. На основі концепції Т-опорних множин запропоновано нове формальне визначення алгоритму розпізнавання. Уведено дерева моделей (ДМ) для задачі класифікації. Показано, що деяка множина моделей, що представляють алгебраїчні системи розпізнавання у вигляді структур дерев класифікації при фіксованих значеннях параметрів перетворюється на конкретну модель класифікації. Запропоновано використання концепції Т-опорних множин для опису дискретних образів.

**У четвертому розділі** Дано числову оцінку впливу процедури перестановки ярусів структури регулярного логічного дерева на його складність для бінарного випадку. Розроблено важливий механізм мінімізації логічних дерев – процедуру перестановки ярусів, структурних блоків у структурі дерева, яка забезпечує досягнення помітного ефекту зменшення складності логічного дерева, причому зі зростанням структурної складності логічного дерева ефективність перестановки ярусів швидко зростає.

**У п'ятому розділі** досліджено питання побудови найскладнішого логічного дерева на основі критерію структурної складності. Дано загальну числову оцінку структурної складності кількості різних міток, функцій, атрибутів найскладнішого логічного дерева в залежності від випадку розташування ярусу злому в його структурі. Розроблено методи знаходження величини подібності для конструкції логічних дерев у задачах мінімізації їх структур. Досліджено питання критерію оптимальності регулярного логічного дерева. Розроблені схеми оптимального розташування змінних у структурі логічного дерева, які в більшості випадків дають оптимальне логічне дерево.

У шостому розділі розроблено комплексний метод побудови дерев класифікації на основі покрової апроксимації масиву початкових даних НВ набором відібраних та оцінених незалежних алгоритмів. Досліджено питання оцінки якості, ефективності, інформативності набору алгоритмів класифікації вершин структури дерева класифікації в схемі методів АДК. Розроблено методи побудови моделей АДК двох типів. Розроблено обмежений метод побудови АДК, який спрямований на побудову лише тих шляхів структури дерева класифікації, де є найбільша кількість помилок класифікації. Досліджено питання збіжності процедури побудови моделей дерев класифікації представлених в роботі методів ЛДК/АДК для умов слабого та сильного розділення класів початкової НВ.

У сьомому розділі розроблено програмну схему обчислення важливості ознак за допомогою введеного в роботу функціонала. Розроблено алгоритмічну схему побудови дерев класифікації, яка забезпечує ефективний механізм програмної побудови фіксованого ЛДК за набором деяких початкових даних. Побудовані моделі дерев класифікації, структури АДК можна застосовувати для оцінки загального стану басейну річки Уж та виявлення ситуації червоної, паводкової зони на основі поточних замірів постів спостережень.

**7. Апробація і повнота викладу нових наукових результатів дисертаційної роботи в опублікованих працях.** Основні результати докторської дисертації Повхана І.Ф. повною мірою відображені у 57 наукових працях, з яких 27 публікацій у фахових наукових виданнях України з технічних наук, 2 монографії, 1 свідоцтво авторського права, 28 публікацій у збірниках матеріалів конференцій. Основні положення дисертації пройшли апробацію автором на міжнародних конференціях.

**8. Ідентичність основних положень автореферату змісту дисертації.** Висвітлені в авторефераті основні положення ідентичні за змістом з викладеними в дисертаційній роботі і не містять інформації, яка не відображена в самій роботі.

Результати та висновки кандидатської дисертації не включено до результатів докторської дисертаційної роботи.

#### **9. Зауваження щодо дисертації:**

1. У першому розділі недостатньо глибоко проаналізовано методи випадкового лісу (випадкових структур ЛДК).

2. У другому розділі при аналізі методів на основі селекції атрибутів варто було б більш детально розібрати типи та випадки формування критеріїв зупинки синтезу моделі дерева класифікації.

3. У третьому розділі для схем задання Т-опорних множин було б доцільно навести приклади їх роботи для кожного з відповідних типів.

4. У четвертому розділі для методу мінімізації, обрізки логічного дерева було б доцільно розглянути приклад нерегулярної структури ЛДК.

5. Із тексту дисертаційної роботи не зрозуміло, які обмеження мають розроблені методи синтезу дерев класифікації.

6. У сьомому розділі роботи в задачі класифікації паводкових станів було б доцільно навести параметри та загальні схеми структур побудованих моделей АДК.

Перелічені зауваження не впливають на загальний високий науковий рівень і практичну цінність дисертаційної роботи.

### **Висновки**

Дисертація Повхана І.Ф. є завершеною науковою працею, в якій викладено нові теоретично обґрунтовані результати, що розв'язують актуальну науково-прикладну проблему розвитку теорії аналізу та синтезу дерев рішень, розроблення моделей, методів, прикладного інструментарію інтелектуального аналізу даних на основі логічних та алгоритмічних дерев класифікації з більшою точністю, зменшеною складністю моделей та підвищеною ефективністю класифікації дискретних об'єктів.

Проведені в дисертаційній роботі дослідження та отримані наукові результати відповідають паспорту спеціальності 05.13.23 – системи та засоби штучного інтелекту. Автореферат повністю відповідає змісту дисертації й описує суть одержаних результатів та висновків і оформлений згідно діючих вимог. Дисертація відповідає вимогам п.п. 9, 10, 12 положення про «Порядок присудження наукових ступенів», а її автор – Повхан Ігор Федорович заслуговує на присудження йому наукового ступеня доктора 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 С.О. Субботіна засвідчую  
Вчений секретар  
Національного університету  
"Запорізька політехніка"



В.В. Кузьмін