

ВІДГУК

офіційного опонента на дисертаційну роботу Повхана Ігоря Федоровича на тему: "Методи та принципи побудови дерев класифікації дискретних об'єктів для інтелектуального аналізу даних", поданої 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 образів, дуже різноманітні та виникають у сучасному світі в усіх сферах економіки та соціального контенту діяльності людини, що приводить до необхідності побудови та дослідження математичних моделей відповідних систем. Станом на зараз не існує універсального підходу до їх розв'язання, запропоновано декілька досить загальних теорій та підходів, що дозволяють вирішувати багато типів задач, але їх прикладні застосування відрізняються досить великою чутливістю до специфіки самої задачі або предметної області застосування. У даний час існує декілька незалежних загальних підходів, концепцій для вирішення задачі класифікації в загальній постановці. В більшості задач прогнозування та класифікації, які використовують неструктуровані дані (наприклад, набори дискретних зображень або текстові масиви), штучна нейронна мережа підбраного типу перевершує за продуктивністю всі інші типи алгоритмів або фреймворків дерев рішень. У протилежному разі (випадку структурованих масивів дискретних даних великого об'єму) в значній мірі перевагу мають методи та алгоритми дерев рішень. Концепція дерев класифікації позбавлена значної частини недоліків класичних методів розпізнавання та дає змогу ефективно працювати в задачах з даними довільних шкал, де інформація задається в природній формі. На сьогоднішній день актуальні різні підходи до побудови систем розпізнавання у вигляді логічних дерев класифікації (ЛДК), причому інтерес до методів розпізнавання, які використовують дерева класифікації викликаний рядом корисних властивостей, якими вони володіють. Отже, актуальною науково-прикладною проблемою в роботі 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, сформульованих у дисертації, їх достовірність і наукова новизна. 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 сформувавши 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, сім розділів, висновки, додатки та список використаних джерел. Загальний обсяг дисертації складає 382 сторінки, з них 67 рисунків, 58 таблиці, 3 додатків та список літератури з 312 найменувань.

3. Дисертаційна робота безпосередньо пов'язана з планами наукових досліджень у Державному вищому навчальному закладі «Ужгородський національний університет» на кафедрах факультету інформаційних технологій у рамках таких науково-дослідних програм, тем, проектів:

- 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 № СВС01008 Норвезького державного фонду із солідарним бюджетом Словацької Республіки в рамках програми SK08 транскордонне співробітництво.

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 та розпізнавання, наведено загальну постановку задачі навчання, виділено основні питання, які виникають на етапі навчання систем розпізнавання. Проаналізовано сучасний стан досліджень концепції дерев рішень, висвітлено проблематику даної концепції, визначено шляхи подальших вдосконалень та перспективних досліджень моделей та методів дерев класифікації. Представлено структурно-логічну схему досліджень.

У другому розділі представлено 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 та надвеликої 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, побудованого за даними навчальної вибірки (НВ).

У третьому розділі розроблено метод T – опорних множин, який полягає у відборі (фіксації) певного набору ознак разом зі своїми значеннями на основі інформації деякої початкової НВ. На основі концепції T – 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 (деревоподібну логічну структуру). Запропоновано використання концепції T – опорних множин для опису дискретних образів.

У четвертому розділі показано, що за підмножиною будь-яких значень набору змінних можна побудувати повне логічне дерево, яке представляє функцію розпізнавання, визначену на всіх наборах. Дано числову оцінку впливу процедури перестановки ярусів структури регулярного логічного дерева на його складність для бінарного випадку, причому ефект перестановки ярусів є значним. Розроблено важливий механізм мінімізації логічних дерев – процедуру перестановки ярусів (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, яке містить у своїй конструкції максимальну кількість різних міток (атрибутів, вершин, функцій). 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, та на основі цього досліджено питання критерію оптимальності регулярного логічного дерева. Розроблені схеми оптимального розташування змінних у структурі логічного дерева, які в більшості випадків дають оптимальне логічне дерево (або близьке до оптимального відносно його структури).

У шостому розділі розроблено комплексний метод побудови дерев класифікації на основі покрової апроксимації масиву початкових даних НВ набором відібраних та оцінених незалежних алгоритмів класифікації та розпізнавання, який дає можливість будувати різнотипні моделі алгоритмічних дерев класифікації (АДК). Досліджено питання оцінки якості (ефективності, інформативності) набору алгоритмів класифікації вершин структури дерева класифікації в схемі методів АДК – відбору критерію розгалуження конструкції дерева класифікації. Розроблено методи побудови моделей АДК двох типів. Розроблено обмежений метод побудови АДК, який спрямований на добудову лише тих шляхів (ярусів) структури дерева класифікації, де є найбільша кількість помилок (усіх типів) класифікації. Досліджено питання збіжності процедури побудови моделей дерев класифікації представлених в роботі методів ЛДК/АДК для умов слабого та сильного розділення класів початкової НВ, наведено відповідні числові оцінки.

У сьомому розділі розроблено програмну схему обчислення важливості ознак (груп ознак) за допомогою введеного в роботу функціонала. Розглядається питання кодування початкової інформації НВ. 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 (на ділянці спостереження) та виявлення ситуації червоної (паводкової) зони на основі поточних замірів постів спостережень. Проведені практичні випробування моделей АДК підтвердили працездатність математичного забезпечення та запропонованих методів й алгоритмів побудови дерев класифікації, розробленого програмного забезпечення.

4. Наукова новизна одержаних результатів. На основі виконаних теоретичних та експериментальних досліджень вирішено важливу науково-прикладну проблему розвитку теорії аналізу та синтезу дерев рішень, розробленню 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 нові результати:

вперше:

- розроблено комплексний метод побудови деревоподібних моделей класифікації, який за рахунок поетапної апроксимації масиву початкових даних набором різноманітних відібраних алгоритмів розпізнавання, забезпечує побудову різнотипних моделей класифікації, їх універсальність та можливість роботи з великими масивами різнотипних даних;

- розроблено метод Т – опорних множин, який шляхом фіксації набору ознак разом зі своїми значеннями на основі початкової вибірки з можливістю оцінки даних опорних множин за допомогою відповідних функціоналів, забезпечує ефективний механізм представлення дискретних об'єктів для

структур логічних дерев класифікації;

- розроблено метод побудови структур алгоритмічних дерев класифікації, які відрізняються модульним принципом побудови моделей, що забезпечує розширення прикладної області застосування, побудову моделей з регульованою точністю класифікації;

- розроблено метод побудови обмежених за складністю структур алгоритмічних дерев класифікації, який за рахунок побудови шляхів конструкції дерева класифікації з найбільшою кількістю помилок забезпечує регулювання складності моделей дерев класифікації;

- розроблено метод знаходження подібності конструкцій логічних дерев, який за рахунок подібних вершин в структурі дерева класифікації забезпечує механізм фінальної обрізки побудованої структури дерева класифікації;

- розроблено метод оцінки впливу процедури обрізки логічного дерева класифікації за рахунок перестановки ярусів, рівнів у конструкції регулярного логічного дерева, що забезпечує зменшення складності його структури;

отримали подальший розвиток:

- методи структур логічних дерев класифікації (випадкових дерев класифікації), які за рахунок схеми виправлення помилок в конструкції логічних дерев шляхом корекції (донавчання) структури дерева класифікації, забезпечують побудову якісно кращих структур дерев класифікації;

- метод оцінки збіжності процедури побудови моделей дерев класифікації (структур дерев алгоритмів) для умов слабого та сильного розділення класів навчальної вибірки за рахунок схем потужності структур дерев класифікації забезпечує отримання параметрів моделі дерева алгоритмів максимальної складності;

удосконалено:

- схеми дерев моделей класифікації, які є зв'язаними графами без циклів, у некінцевих вершинах яких знаходяться фіксовані моделі, ребра нумеруються значеннями предикатів цих моделей, що забезпечує спрощення побудови нових класифікаторів за рахунок використання модульного принципу.

6. Практична значущість одержаних результатів в тому, що розроблені методи, моделі та засоби синтезу структур дерев класифікації дають змогу: будувати різнотипні моделі дерев класифікації, підвищити їх універсальність, працювати з великими масивами різнотипних даних; організувати ефективний механізм представлення дискретних об'єктів для структур логічних дерев класифікації; розширити прикладну область застосування, будувати моделі з регульованою точністю класифікації; реалізувати механізм фінальної обрізки побудованої структури дерева класифікації; забезпечити зменшення складності структур логічних дерев; будувати якісно кращі структури дерев рішень; отримання параметрів моделі дерева алгоритмів максимальної складності;

Розроблені в дисертаційному дослідженні концепції, методи та моделі синтезу систем розпізнавання доведено до практичної реалізації в інформаційних системах, зокрема для аналізу та класифікації екологічних ситуації в басейнах річок Закарпатського регіону (акт впровадження від

07.09.2011). Ряд побудованих моделей і схем АДК у межах цього дослідження використано ТОВ «ІНФОСФЕРА» (акт впровадження від 04.05.2019) та ТОВ «Медіа - Сервіс» (акт впровадження від 11.09.2020) м. Ужгород, Управлінням економічного розвитку міста Ужгородської міської ради (акт впровадження від 04.09.2020), Закарпатською обласною громадською організацією «Патріотичний оборонний – спортивний центр Вітязь» (акт впровадження від 18.05.2012).

7. Апробація і повнота викладу нових наукових результатів дисертаційної роботи в опублікованих працях. Основні результати докторської дисертації Повхана І.Ф. відображено в 57 наукових публікаціях, серед яких 2 монографії, 27 статей у фахових виданнях України з технічних наук, 28 публікацій у матеріалах і тезах конференцій, 1 свідоцтво про реєстрацію авторського права на твір (комп'ютерна програма). Основні положення дисертації пройшли апробацію автором на міжнародних конференціях.

Ідентичність основних положень дисертації 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.

8. Зауваження щодо дисертації:

1. При аналізі фреймворків C4.5 та C5.0 на початку роботи бажано більш детально розглянути концептуальну різницю в плані генерації базових критеріїв схеми дерева класифікації.
2. В третьому розділі з основного означення T – опорної множини впливає існування більше ніж чотирьох базових класів систем T – опорних множин, про що треба сказати більш детально.
3. У шостому розділі не показані основні обмеження які накладаються на набори незалежних алгоритмів класифікації в схемі генерації моделей алгоритмічного дерева класифікації.
4. Для обмеженого методу алгоритмічного дерева класифікації з шостого розділу роботи доцільно би було показати вплив параметру препрунінгу (Z) на загальну складність моделі дерева алгоритмів.
5. Доцільно було би порівняти метод алгоритмічне дерево класифікації з фреймворками LightGBM та XGBoost на даних задачі класифікації паводкових явищ представленої в роботі в сьомому розділі дослідження.
6. По тексту дисертації зустрічаються окремі термінологічні і стилістичні неточності.

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 у дисертаційній роботі і оформлений згідно діючих вимог. Дисертація відповідає вимогам зокрема, п.п. 9, 10, 12 положення про «Порядок присудження наукових ступенів», а її автор, Повхан Ігор Федорович, заслуговує присудження наукового ступеня доктора технічних наук за спеціальністю 05.13.23 – системи та засоби штучного інтелекту.

Офіційний опонент:


Завідувач кафедри інформатики та

комп'ютерних наук

Херсонського національного технічного

університету, доктор технічних наук,

професор

 В. І. Литвиненко

Підпис Литвиненка Володимира

Івановича засвідчую.

Начальник відділу кадрів ХНТУ



М.В. Танська