

ВІДГУК

офіційного опонента про дисертацію

Повхана Ігоря Федоровича «Методи та принципи побудови дерев класифікації дискретних об'єктів для інтелектуального аналізу даних», представленої на здобуття наукового ступеня доктора технічних наук за спеціальністю 05.13.23 – системи та засоби штучного інтелекту

1. Актуальність теми. Дисертаційна робота є завершеним науковим дослідженням і присвячена вирішенню актуальної науково-прикладної проблеми розвитку методів і алгоритмів аналізу та синтезу дерев рішень, розробки програмних засобів інтелектуального аналізу даних на основі підходу логічних та алгоритмічних дерев класифікації для підвищення ефективності процесів класифікації в інформаційних системах. Так, концепція дерев класифікації має певні переваги над класичними підходами класифікації. На сьогоднішній день актуальні різні підходи до побудови систем розпізнавання у вигляді логічних дерев класифікації. Це пояснюється тим, що дерева класифікації мають ряд корисних властивостей.

За результатами роботи проведено експериментальне моделювання, апробація розроблених моделей та методів на практичних задачах. Визначення ефективності розроблених в роботі методів побудови дерев алгоритмів здійснено на основі інтегральних показників якості моделей. У роботі розроблена, теоретично обґрунтована нова організація побудови моделей алгоритмів розпізнавання та класифікації дискретних об'єктів великих масивів різнотипної інформації. Розроблено програмну систему для розв'язку задач розпізнавання образів. Розроблений в дисертаційному дослідженні підхід (на основі концепції структур дерев алгоритмів та модульного принципу) дає можливість розв'язувати прикладні задачі аналізу даних, розпізнавання образів, розпізнавання 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 коректною постановкою задач та їх подальшим теоретичним викладенням і співпадінням із результатами практичних досліджень та апробацією на конференціях і семінарах.

3. Структура та зміст дисертації.

Дисертаційна робота має чітку структуру і складається з семи розділів.

У першому розділі проведено аналіз деревоподібних моделей класифікації та розпізнавання. Наведено загальну постановку задачі навчання. Проаналізовано сучасний стан досліджень методів і алгоритмів дерев рішень, 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 використання об'єктів вибірки в своїй структурі. Також показано відсутність ефективних методів та алгоритмів, які би дозволили будувати конструкції логічних дерев на основі масивів даних великої та надвеликої розмірності. Розглянуто окремих клас дерев класифікації – випадкові ЛДК, які мають як переваги (програмна простота побудови дерева, зменшення часу загальної генерації (відбору) вершин ЛДК, можливість оцінки та вибору найвідповіднішого ЛДК із множини побудованих), так і недоліки (неоптимальність структури, апаратні витрати на генерацію та зберігання неоптимального ЛДК). Розроблено алгоритми 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 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, з можливістю наступної оцінки цих опорних множин за допомогою відповідних функціоналів. Розроблено основні способи задання систем T – опорних множин. На основі T – опорних множин запропоновано нове формальне визначення алгоритму розпізнавання (класифікації). Досліджено, що найбільший інтерес викликають саме прості системи T – опорних множин. Уведено дерева моделей (ДМ) для задачі

класифікації, які є зв'язаним графом без циклів, в некінцевих вершинах якого знаходяться фіксовані моделі, а ребра нумеруються значеннями предикатів (узагальнених ознак) з цих моделей. 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 використовувати T – опорні множини для опису дискретних образів, на основі чого представлено нові означення інформативності відносно класу та об'єкту класифікації. Також введено поняття $2T$ – опорних множин, які можна використовувати в якості ознак, що характеризують деякий образ.

У четвертому розділі показано, що за підмножиною будь-яких значень набору змінних можна побудувати повне логічне дерево, яке є функцією розпізнавання, яка визначена на всіх наборах, що забезпечує можливість застосування логічних дерев у розпізнаванні образів. Показано, що принциповою задачею є отримання мінімальної (оптимальної) структури ЛДК відносно початкової НВ. 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 (блоків) у структурі дерева, яка забезпечує досягнення ефекту зменшення складності логічного дерева, причому із зростанням структурної складності логічного дерева ефективність перестановки ярусів швидко зростає.

У п'ятому розділі досліджено побудову 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. Розроблено на основі структури логічного дерева мінімальної складності схему мінімізації логічної функції, яка базується на побудові мінімального логічного дерева та мінімізації функції.

У шостому розділі розроблено комплексний метод побудови дерев

класифікації на основі покрокової апроксимації масиву початкових даних НВ набором, відібраних та оцінених незалежних алгоритмів класифікації та розпізнавання. Досліджено оцінку якості (ефективності, інформативності) набору алгоритмів класифікації. Розроблено методи побудови моделей АДК двох типів, причому отримані дерева класифікації складаються з різних алгоритмів розпізнавання. 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 збіжність 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 цьому розділі розроблено алгоритми обчислення важливості ознак (груп ознак) за допомогою введеного в роботі функціонала. Розроблено алгоритми побудови дерев класифікації. Побудовані моделі дерев класифікації (структури АДК) застосовані для оцінки загального стану басейну річки Уж (на ділянці спостереження) та виявлення ситуації червоної (паводкової) зони на основі поточних замірів постів спостережень. Проведені практичні випробовування моделей АДК підтвердили адекватність розроблених методів й алгоритмів побудови дерев класифікації. Це дає змогу рекомендувати даний підхід та його програмну реалізацію для ряду прикладних задач класифікації та розпізнавання.

3. Наукова новизна одержаних результатів. На основі виконаних теоретичних та експериментальних досліджень вирішено важливу 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 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 T – 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.

4. Зв'язок з науковими програмами, планами, темами. Дисертаційна робота виконана у Державному вищому начальному закладі «Ужгородський національний університет» на кафедрах факультету інформаційних технологій відповідно до плану наукових досліджень 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.

5. Практичне значення одержаних результатів полягає в тому, що на основі запропонованих методів розроблено:

- алгоритми обчислення важливості ознак (груп ознак);
- алгоритми побудови дерев класифікації;
- програмну систему – DeTree для побудови обмежених моделей дерев алгоритмів, що зменшило час синтезу моделі АДК на 15-20% і зменшило структурну складність логічного дерева в процедурі обрізки ЛДК на основі схеми перестановки ярусів в 2.5 – 3 рази.

6. Апробація і повнота викладу нових наукових результатів дисертаційної роботи в опублікованих працях. Результати дисертаційного дослідження повною мірою відображені у 57 наукових працях, у тому числі 2 монографіях, 27 статтях у фахових наукових виданнях України з технічних

наук, 8 із них індексуються у міжнародних наукометричних базах, 1 авторське свідоцтво, 28 публікацій в збірниках матеріалів конференцій.

Основні положення дисертації пройшли апробацію автором на міжнародних науково-практичних конференціях. 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.

7. Зауваження щодо дисертації:

1. У першому розділі доцільно було б приділити більше уваги аналізу сучасних методів і алгоритмів дерев класифікації, виділити критерії порівняння та здійснити їх порівняння.

2. Автор виділяє переваги методів і алгоритмів логічних дерев класифікації порівняно з штучними нейронними мережами. Так, дійсно, логічні дерева класифікації дають змогу представити процес класифікації в явному вигляді на відміну від штучних нейронних мереж. Але, на даний час, наприклад, найкращими класифікаторами зображень без необхідності виділення ознак є згорткові нейронні мережі, які можна використати не тільки для класифікації але для генерування зображень (генеративно-змагальні мережі).

3. У четвертому розділі не зрозуміло, чи можливе застосування процедури перестановки ярусів, схеми обрізки для конструкції логічних дерев інших типів (класів).

4. Перед побудовою логічних дерев класифікації доцільно було б провести ранжування ознак об'єктів, наприклад методом головних компонент. Це дало змогу зменшити складність дерев класифікації.

5. Побудову алгоритмічних дерев класифікації доцільно було б почати із оцінки обчислювальної складності алгоритмів.

6. В задачі побудови моделей класифікації ситуаційних станів басейну річки Уж доцільно було б провести порівняння побудованих структур алгоритмічних дерев класифікації із схемами випадкового лісу.

7. В питанні збіжності процедури побудови структури для умов слабого і сильного розділення класів початкової навчальної вибірки доцільно провести оцінку і для випадку моделі алгоритмічних дерев класифікації другого типу.

8. Для представлення розробленої програмної системи необхідно було б привести його узагальнену структуру, опис основних модулів і здійснити порівняння із відомими на ринку програмними системами.

9. По тексту дисертації зустрічаються граматичні та орфографічні помилки, термінологічні 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, 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.

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 комп'ютерної інженерії

Західноукраїнського національного університету

д.т.н., професор

Олег БЕРЕЗЬКИЙ

Підпис *О. Березький*
Завіряю: *А. Місякович*
НАЧАЛЬНИК ВІДДІЛУ
ОРГАНІЗАЦІЇ ІНФОРМАЦІЙНОГО ТА
ДОКУМЕНТАЛЬНОГО ЗАБЕЗПЕЧЕННЯ

