

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ЛЬВІВСЬКА ПОЛІТЕХНІКА»

**ПОВХАН ІГОР ФЕДОРОВИЧ**



УДК 004.8: 004.89: 519.7

**МЕТОДИ ТА ПРИНЦИПИ ПОБУДОВИ ДЕРЕВ КЛАСИФІКАЦІЇ  
ДИСКРЕТНИХ ОБ'ЄКТІВ ДЛЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ**

05.13.23 – Системи та засоби штучного інтелекту

**АВТОРЕФЕРАТ**

дисертації на здобуття наукового ступеня  
доктора технічних наук

**Львів – 2021**

**Дисертацією є рукопис.**

**Робота виконана у ДВНЗ «Ужгородський національний університет»  
Міністерства освіти і науки України**

**Науковий консультант:**

доктор технічних наук, професор  
**Гече Федір Елемирович,**  
ДВНЗ “Ужгородський національний університет”  
завідувач кафедри кібернетики і прикладної математики

**Офіційні опоненти:**

доктор технічних наук, професор  
**Березький Олег Миколайович,**  
Західноукраїнський національний університет  
завідувач кафедри комп'ютерної інженерії

доктор технічних наук, професор  
**Литвиненко Володимир Іванович,**  
Херсонський національний технічний університет  
завідувач кафедри інформатики і комп'ютерних наук

доктор технічних наук, професор  
**Субботін Сергій Олександрович,**  
Запорізький національний технічний університет  
завідувач кафедри програмних засобів

Захист відбудеться «06» травня 2021 р. о 14:00 год. на засіданні спеціалізованої вченої ради Д 35.052.14 у Національному університеті “Львівська політехніка” за адресою: 79013, м. Львів, вул. С. Бандери, 12, ауд. 226, головний навчальний корпус.

З дисертацією можна ознайомитись у бібліотеці Національного університету “Львівська політехніка” за адресою: 79013, м. Львів, вул. Професорська, 1.

Автореферат розісланий «31» Березня 2021 р.

**Учений секретар  
спеціалізованої вченої ради,  
к.т.н., доцент**

**А. Є. Батюк**

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

**Актуальність теми.** 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. Багато теоретичних результатів отримано для спеціальних випадків та підзадач, причому слід відмітити, що вузьким місцем вдалих реальних систем розпізнавання залишається необхідність виконання величезного об'єму обчислень та орієнтація на потужний апаратний інструментарій. Проте, велика кількість прикладних задач в різних галузях природознавства, наприклад, в геології, геофізиці, геохімії, медицині, соціології, археології, біології та інше, де вирішуються задачі класифікації з використанням програмних та апаратних систем, визначає інтенсивність та актуальність такого напрямку досліджень.

Нейромережева концепція розпізнавання – не дивлячись на значні переваги, має, тим не менше, істотні недоліки, які обмежують галузь її застосування. Нейронні мережі дають змогу знаходити відповідні субоптимальні розв'язки, що є проблемою в задачах з вимогою високої точності моделі, що будується. Загальна схема функціонування зводиться до принципу чорної скрині, що недопустимо за умови аналізу причини прийняття того чи іншого рішення. Значні апаратні та часові витрати інформаційної системи на процес навчання моделі не компенсуються швидкістю фінальної класифікації. Обмеження на формат вхідних даних, яке зводиться до числової шкали для методів нейромереж, накладає додаткове принципове обмеження щодо спектру можливих прикладних задач. Тому слід визнати, що клас реальних задач, які підпадають під ці обмеження достатньо широк.

Так, концепція дерев класифікації позбавлена значної частини наведених вище недоліків 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 (СР) у вигляді логічних дерев класифікації (ЛДК), причому інтерес до методів розпізнавання, які використовують ЛДК, викликаний рядом корисних властивостей, якими вони володіють. З одного боку, складність класу функцій розпізнавання (ФР) у вигляді моделей ЛДК, при визначених умовах, не перевищує складності класу лінійних функцій розпізнавання (простішого з відомих). З іншого боку, ФР у вигляді дерев класифікації дозволяють виділити в процесі класифікації як причинно-наслідкові зв'язки та однозначно врахувати їх у подальшому, так і фактори випадковості або невизначеності, тобто врахувати одночасно і функціональні, і стохастичні відношення між властивостями та поведінкою всієї системи, причому відомо, що процес класифікації нових, таких що до сих пір не зустрічалися об'єктів світу багатьох тварин і людей (за виключенням об'єктів, інформація про які передається генетичним шляхом, наслідковим, а також в деяких інших випадках), відбувається за так званим логічним деревом рішень.

Зафіксуємо, що 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 концепції дерев рішень.

У даний час існує декілька незалежних загальних підходів, концепцій для вирішення задачі класифікації в загальній постановці, причому розробку різних концепцій, підходів, методів, моделей, які охоплюють загальну проблематику теорії штучного інтелекту та інформаційних систем проводили і проводять відомі як вітчизняні так і закордонні вчені, науковці, галузеві спеціалісти, серед яких можна виділити наступних: М.А. Айзерман, Ю.І. Журавльов, А.Г. Аркадьєв, А.Ш. Блох, В.Н. Вапник, А.Н. Червоненкіс, В.І. Васильєв, Б.А. Головкин, І.Б. Гуревич, Ю.А. Зуєв, А.І. Кондратьєв, Г.С. Лбов, В.А. Орлов, А.Б. Глаз, А.А. Алексанян, В.А. Євстегнеєв, В.Є. Котов, Н.Г. Загоруйко, Л.А. Растрігін, Р.Х. Еренштейн, А.А. Барсегян, Н.А. Лагуновский, С.В. Абраменко, Д.М. Абрамов, Є.В. Бодяньський, В.П. Машталір, Ю.А. Василенко, С.А. Суботин, Д. Квінлан, Г. Денг, Б. Камінські, К. Карімі, Т. Хасті, Д. Стоун, Р. Олсен, Р. Ріверс, Л. Хайфіл, П. Гертс, Т. Хортон, Л. Брейман, М. Міакава, Д. Фрідман, К. Верхаген, Р. Дейн, Ф. Грун, П. Вебек, А. Ахо, Н. Вірт, В. Ліпські, А.Фор, П. Вогоф, Г. Віттен, Є. Франк, Д. Макленнен, Л. Шапіро, Д. Стокман, Д. Форсайт, Р. Дуда, П. Харт, Д. Кнут, У. Претт, М. Хетч, Ш. Чен, Д. Аміт, Д. Геман, К. Вілдер, П. Лавракс, А. Фішер, Р. Тамасія, Р. Хантер та інші.

Усі ці підходи в теорії розпізнавання мають свої переваги і недоліки та утворюють єдиний інструментарій розв'язку прикладних задач теорії штучного інтелекту. Так, цілісно пропрацьованим з математичної точки зору є класичний алгебраїчний підхід, розроблений Ю.І. Журавльовим. Цей напрямок розвитку теорії розпізнавання зв'язаний з побудовою моделей алгоритмів класифікації та вибором у рамках моделі оптимального за якістю алгоритму розпізнавання, причому значна увага в межах дослідження приділена алгоритмам обчислення оцінок, на основі чого вводиться поняття  $T$  – опорної множини, демонструється зв'язок  $T$  – опорних множини та дерев класифікації. Центральну увагу в дослідженні приділено актуальній концепції дерев рішень. Відомо, що схема класифікації, яка задається довільним підходом, методом, алгоритмом дерева класифікації має деревоподібну логічну структуру, причому структура логічного дерева складається з вершин, які групуються за ярусами та побудовані, відібрані на певному кроці побудови моделі дерева класифікації, а головна особливість деревоподібних СР полягає в тому, що важливість окремих ознак (групи ознак чи їх наборів) визначається відносно функції, яка задає розбиття об'єктів на класи. Важливим напрямком досліджень структур ЛДК залишаються питання стосовно генерації дерев рішень у разі малоінформативних ознак та актуальне питання теорії дерев класифікації – питання можливої побудови всіх варіантів логічних дерев, які відповідають початковій навчальній виборці (НВ) та відбору мінімального за глибиною, структурною складністю дерева класифікації.

**Отже, актуальною науково-прикладною проблемою 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.**

**Зв'язок роботи 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, проєктів:

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

**Мета і задачі дослідження.** Метою дисертаційної роботи є розроблення нової теоретичної основи побудови моделей розпізнавання у вигляді алгоритмічних дерев класифікацій (структур АДК) та їх застосування для автоматизації розробки систем розпізнавання образів.

Досягнення поставленої мети передбачало розв'язання таких задач:

- аналіз властивостей різних методів, алгоритмів розпізнавання (випадкових логічних дерев у тому числі), заснованих на схемах логічних дерев класифікації;

- аналіз та класифікація моделей розпізнавання образів у вигляді дерев рішень – різнотипних структур логічних та алгоритмічних дерев класифікації;

- виявлення та встановлення зв'язків опорних множин з структурами дерев класифікації в задачі представлення дискретних об'єктів;

- розроблення методів знаходження подібності конструкції логічних дерев класифікації в задачах мінімізації їх структур та на основі цього дослідження питання критерію оптимальності регулярного логічного дерева;

- розроблення схем оптимального розташування змінних (атрибутів) у структурі логічного дерева, які найчастіше дають оптимальне логічне дерево відносно його структури.

- розроблення загальної схеми обчислення складності структур класу регулярних логічних дерев для бінарного випадку;

- розроблення методів побудови моделей дерев алгоритмів на основі покрокової апроксимації масиву початкових даних навчальної вибірки набором оцінених незалежних алгоритмів класифікації та розпізнавання;

- встановлення питання збіжності процедури побудови моделей дерев класифікації методів логічних та алгоритмічних дерев для умов слабкого та сильного розділення класів початкової навчальної вибірки;

- розроблення програмного інструментарію побудови структур логічних та алгоритмічних дерев класифікації з метою розв'язку прикладних задач класифікації образів.

**Об'єктом дослідження** є процеси синтезу структур логічних дерев класифікації різних типів та схем.

**Предмет дослідження** – методи, алгоритми, моделі та інструментальні засоби побудови логічних дерев класифікації.

**Методи дослідження.** При проведенні досліджень та вирішенні вище зазначених задач використовувались методи: методи та математичний апарат теорії розпізнавання образів, розпізнавання зображень, концепція організації процесу розпізнавання на основі дерева класифікації (дерева рішень), математичної статистики, нечітких множин, теорії графів, алгебри – логіки, багатозначної логіки, логіко-комбінаторні методи, методи теорії алгоритмів, методи теорії дискретних функцій, методи структурного та об'єктно-орієнтованого програмування для розробки програмного забезпечення інформаційних систем, використовувались програмні системи для розв'язку прикладних задач – ОРІОН III, відкриті ООБ дерев рішень LightGBM, XGBoost.

**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;

**набули подальшого розвитку:**

- 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.

**Практичне значення одержаних результатів** полягає в тому, що розроблені методи, алгоритми та інструментальні засоби 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.

**Реалізація і впровадження результатів роботи.** Розроблені в дисертаційному дослідженні концепції, методи та моделі синтезу СР доведено до практичної реалізації в інформаційних системах, зокрема для аналізу та класифікації екологічних ситуацій 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).

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» (акт впровадження від 31.08.2020).

**Особистий внесок здобувача.** З набору наукових праць, опублікованих автором у співавторстві, в дисертаційному дослідженні використані ідеї та положення, які є результатом індивідуальної праці автора. Праці опубліковані одноосібно: [1,2] – методи мінімізації регулярних логічних конструкцій; [3-7,33] – методи побудови структур ЛДК/АДК, обмежених дерев класифікації в тому числі; [8-13] – схеми побудови та мінімізації логічних дерев; [14] – метод дерева алгоритмів другого типу; [15-20,30] – схеми синтезу дискретних зображень, схеми побудови структур ЛДК; [22] – схема випадкових логічних дерев класифікації; [23,24,29] – схеми геометричної апроксимації; [28] – метод синтезу узагальнених ознак в структурах АДК; [41] – схема усунення помилок розпізнавання; [42-47] – загальна проблематика методів ЛДК/АДК. У наукових роботах, написаних у співавторстві, автору належать: загальна постановка задачі, вибір або розробка методів та алгоритмічних схем їх розв'язку, ідеї використання концепції апроксимації даних навчальної вибірки набором незалежних автономних алгоритмів класифікації. Зокрема, у публікаціях здобувачу належать: [25,34,38] – постановка задачі, методика та алгоритмічні схеми обрахунку важливості дискретних ознак, наборів та їх сполучень; [26,27] – схеми оцінки складності структур логічних дерев розпізнавання, метод мінімізації деревоподібних логічних структур та схема оцінки ефекту такої мінімізації, оцінка стійкості максимального за структурною складністю логічного дерева відносно процедури перестановки ярусів; [35,36] – модульний метод побудови схем (структур дерев класифікації) класифікації та розпізнавання, підхід представлення великих масивів даних за допомогою відповідних структур (ЛДК/АДК), загальний метод побудови моделей АДК; [37,39] – методика представлення правил класифікації в задачах розпізнавання образів у вигляді схем – агентів; [21] – алгоритми та загальні схеми побудови моделей дерев класифікації, алгоритм донавчання та виправлення помилок класифікації в структурі ЛДК; [48] – метод та схема алгоритму генерації наборів УЗ в структурах АДК, набір синтезованих моделей АДК; [31] – загальна концепція алгоритмічних дерев, метод побудови АДК (типу I), набір побудованих моделей АДК; [32] – побудова та порівняльний аналіз запропонованих моделей.

**Апробація результатів дисертації.** Основні теоретичні та практичні аспекти та результати дисертаційної роботи представлено та обговорено на таких науково-практичних конференціях та семінарах: IV Міжнародній науковій конференції «Сучасні тенденції інформаційних технологій», м. Прага (Чеська Республіка), 2008; XV Міжнародній науково-практичній конференції «Інновації в навчальному процесі вищих навчальних закладів», м. Сніна (Словацька Республіка), 2008; IV Міжнародній школі-семінарі «Теорія прийняття рішень», м. Ужгород, 2008; II Міжнародній науковій конференції молодих вчених Інституту електронної фізики Національної академії наук України «ІЕФ-2011», м. Ужгород, 2011; III Міжнародній науковій конференції «Математическое моделирование, оптимизация и информационные технологии», м. Кішеневу (Республіка Молдова), 2012; Міжнародному науковому семінарі «Innovative Methods in Education and Research – InovEduc»,



м. Прага (Чеська Республіка), 2015; Міжнародній науково-практичній конференції «Data Stream Mining & Processing – DSMP», м. Львів, 2016, 2020; Міжнародній школі-семінарі «Communications in Computer and Information Science – CCIS», м. Львів, 2000; III Міжнародній науково-практичній конференції «Computer Modeling and Intelligent Systems – CMIS», м. Запоріжжя, 2020; III Міжнародній науково-практичній конференції «Сучасні проблеми математичного моделювання, автоматизованого керування та інформаційних технологій – MCIT», м. Рівне, 2019; II Міжнародній науково-практичній конференції «Priority directions of science development», м. Львів, 2019; Міжнародній науково-практичній конференції «Інформаційні технології та комп'ютерне моделювання ІТКМ», м. Івано-Франківськ, 2019; Міжнародній науково-практичній конференції «Science, Engineering and Technology: Global and Current Trends», м. Прага (Чеська Республіка), 2019, 2020; X Міжнародній науково-практичній конференції «Математичні методи, моделі та інформаційні технології в економіці», м. Чернівці, 2019; Міжнародній науково-практичній конференції «Цифрова економіка та інформаційні технології», м. Київ, 2020; X Міжнародній науково-практичній конференції «Комплексне забезпечення якості технологічних процесів та систем», м. Чернігів, 2020; XX Міжнародній науково-технічній конференції «Проблеми інформатики та моделювання», м. Харків – м. Одеса, 2020; V Всеукраїнській науково-практичній конференції «Перспективні напрямки сучасної електроніки, інформаційних і комп'ютерних систем», м. Дніпро, 2020; III Всеукраїнській науково-технічній конференції «Комп'ютерні технології: інновації, проблеми, рішення», м. Житомир, 2020.

**Публікації за темою дисертації.** 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.

**Структура та обсяг дисертації.** Дисертаційна робота має класичну структуру та складається зі вступу, семи розділів, висновків, списку використаних джерел із 317 найменувань та 3 додатків. Обсяг дисертації становить 382 сторінки, у тому числі 276 сторінок основного тексту, робота містить 67 рисунків та 58 таблиць.

## **ОСНОВНИЙ ЗМІСТ РОБОТИ**

**У вступі** обґрунтовано актуальність теми дисертаційної роботи, зафіксовано мету та задачі, об'єкт, предмет і методи дослідження. Наведено наукову новизну представленого дослідження та практичне значення одержаних результатів, загальну кількість публікацій за темою роботи у фахових виданнях України з технічних наук і тезах та матеріалах конференцій, виокремлено особистий внесок здобувача.

**У першому розділі** показано, що всебічний аналіз та дослідження деревоподібних моделей класифікації та розпізнавання залишається актуальною задачею теорії штучного інтелекту, хоча головна увага зміщена в бік нейромережевого розпізнавання.

Аргументовано, що представлення масивів даних (дискретної інформації) великого об'єму у вигляді структури логічного дерева має свої суттєві переваги щодо економічного опису даних та ефективних механізмів роботи з ними, а галузь застосування концепції дерев класифікації може бути зведена до задач опису структур

даних, задач розпізнання та класифікації, задач регресії. Показано, що важливою особливістю дерев класифікації є гнучкість, здатність структури ЛДК послідовно враховувати та досліджувати ефект впливу окремих змінних, атрибутів структури.

У межах дослідження ставиться задача аналізу та синтезу таких методів, моделей розпізнавання, які би давали можливість в процесі навчання побудувати можливість просту деревоподібну схему розпізнавання (схему у вигляді структур ЛДК або АДК), яка забезпечує необхідну ефективність та складність системи розпізнавання.

Підкреслено, що при побудові структур дерев класифікації центральними питаннями залишаються питання вибору критерію розгалуження, критерію зупинки навчання та критерію відкидання гілок логічного дерева. Не дивлячись на певну проблематику, яка виникає при побудові та використанні концепції дерев класифікації, в загальному сенсі вона залишається потужним інструментом теорії штучного інтелекту.

У розділі наведено загальну постановку задачі машинного навчання в довільних шкалах, яка зводиться до навчання СР обрахунку ФР (ФР визначена на початковій множині та приймає скінчену кількість значень), причому на даному етапі СР приймає послідовність навчальних пар НВ та на основі цієї інформації будує схему обчислення ФР, або її наближення.

Показаний загальний еволюційний шлях методів та схем теорії дерев рішень. На основі аналізу сучасного стану методів логічних дерев класифікації (структур ЛДК) встановлено, що існує значна кількість схем, які реалізують концепцію дерев рішень, але найбільшого вживання та розповсюдження отримали класи алгоритмів CART та C4.5/C5.0. На основі досліджень визначено загальну проблематику (п'ять проблемних питань) теорії дерев класифікації, яким і присвячено дисертаційне дослідження.

Виділено основні питання, які виникають на етапі навчання СР – економія пам'яті, швидкість навчання системи, питання побудови оптимальної (за фіксованими параметрами) схеми для обчислення ФР. Аргументовано, що можливим раціональним розвитком методів та підходів ЛДК може бути використання ідеї алгоритмічних дерев класифікації. Запропоновано побудову моделей дерев класифікації на основі різнотипних алгоритмів розпізнавання, використовуючи їх в якості вершин структури АДК, використовуючи модульну концепцію, будувати новий алгоритм розпізнавання на основі відомих.

У розділі висвітлено проблему та загальні задачі дисертаційної роботи, представлено напрямки та структурно-логічну схему досліджень.

У **другому розділі** представлено методи дерев класифікації, які забезпечують покриття масиву даних навчальної інформації структурою ЛДК (Рис. 1) за рахунок фіксації об'єктів вибірки в своїй структурі та методи на основі селекції елементарних ознак. Навчальна інформація задається на основі НВ наступного вигляду:

$$\left( (x_1, f_R(x_1)), (x_2, f_R(x_2)), \dots, (x_m, f_R(x_m)) \right).$$

Ці підходи зберігання інформації у вигляді структур ЛДК забезпечують механізм донавчання, розширення структури та виправлення помилок, а застосування ЛДК у задачах розпізнавання дозволяє просто та компактно описувати ФР. Зафіксовано відсутність актуальних методів та алгоритмів, які би дозволили будувати ефективні

конструкції логічних дерев на основі масивів даних НВ великої та надвеликої розмірності.

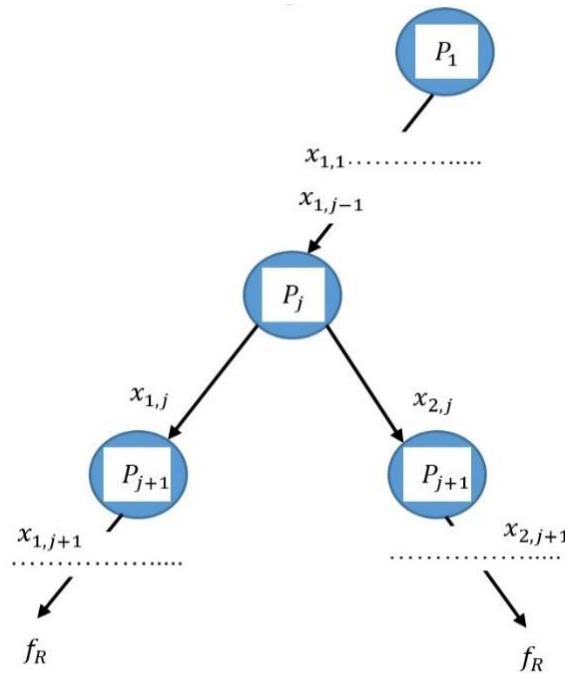


Рисунок 1 – Загальна схема структури ЛДК

У роботі досліджено окремий клас дерев класифікації – випадкові структури ЛДК, які мають суттєві переваги – програмна простота побудови дерева, зменшення часу загальної генерації (відбору) вершин ЛДК, можливість оцінки та вибору найвідповіднішого ЛДК із множини побудованих. Розроблена схема побудови випадкових ЛДК дозволяє генерувати цілі набори дерев класифікації різної структури, складності. Для зберігання фактичної структури побудованого випадкового ЛДК використовується лише програмний ідентифікатор (паспорт) логічного дерева, який містить лише фактичну послідовність вершин (атрибутів у цьому дереві). Такий підхід дозволяє організувати ресурсно-економний спосіб представлення складних структур даних.

Інформація, що вказує тільки на порядок розміщення ознак в логічному дереві, називається схемою ЛДК, а кількість ознак, що входять в цю схему – її порядком (потужністю). Структура ЛДК представляє собою зв'язний граф без циклів, у некінцевих вершинах якого знаходяться ознаки (компоненти об'єктів  $x_1, \dots, x_m$ ):

$$x_1 = (x_1^1, x_2^1, \dots, x_n^1), \dots, x_m = (x_1^m, x_2^m, \dots, x_n^m); P_i(x) = x_i^j, (i = 1, \dots, n), (j = 1, \dots, m).$$

Схема усунення помилок розпізнавання в структурі ЛДК полягає в наступному:

Крок 1. У парі  $(x_{m+1}, f_R(x_{m+1}))$  об'єкту  $x_{m+1} = (x_1^{m+1}, \dots, x_n^{m+1})$  відповідає свій шлях  $a_1, a_2, \dots, a_r; (r \in \{1, \dots, n\})$ , який закінчується значенням  $f_R(x_{m+1})$ .

$$P_{i_1} = q(a_1), a_2 = (a_1)_{x_{i_1}^h}, P_{i_2} = q(a_2), a_3 = (a_2)_{x_{i_2}^h}, \dots$$

$$\dots, P_{i_{r-1}} = q(a_{r-1}), a_r = (a_{r-1})_{x_{i_{r-1}}^h}.$$

Тут  $f_R(x_h) = q(a_r), (h = 1, 2, \dots, m), 1 \leq r \leq n$ , відповідно  $i_1, i_2, \dots, i_{r-1}$  номери ознак, що знаходяться в вершинах  $a_1, a_2, \dots, a_{r-1}$  та  $f_R(x_{m+1}) = q(a_r)$  (такий запис означає, що значення функції  $f_R(x_{m+1})$  не співпадає зі значенням  $f_R(x_h)$ , що знаходиться в некінцевій вершині  $a_r$  тобто виникла помилка розпізнавання). Запам'ятовуємо, що  $f_R(x_h) = q(a_r)$  та після цього проробляємо наступну процедуру:

$P_{i_r} = q(a_r)$ ,  $a_{r+1} = (a_r)_{x_{i_r}^{m+1}}$  (це означає, що з вершини  $a_r$  виходить стрілка  $x_{i_r}^{m+1}$ , відповідна значенню ознаки  $P_{i_r}$  (об'єкту  $x_{m+1}$ ) та входить у наступну за нею вершину  $a_{r+1}$ ).

Крок 2. На наступному кроці будемо мати:

$$P_{i_{r+1}} = q(a_{r+1}), a_{r+2} = (a_{r+1})_{x_{i_{r+1}}^{m+1}}, \dots, P_{i_t} = q(a_t), a_{t+1} = (a_t)_{x_{i_t}^{m+1}},$$

$$P_{i_{t+1}} = q(a_{t+1}), a_{t+2} = (a_{t+1})_{x_{i_{t+1}}^{m+1}}, \dots, P_{i_n} = q(a_n), a_{n+1} = (a_n)_{x_{i_n}^{m+1}}.$$

Тут  $f_R(x) = q(a_{n+1})$ ,  $i_1, i_{r+1}, \dots, i_t, i_{t+1}, \dots, i_n$  – номери ознак, що знаходяться в вершинах ЛДК  $a_r, a_{r+1}, \dots, a_t, a_{t+1}, \dots, a_n$ ,  $r \leq t \leq n - 1$ .

Крок 3. На наступному етапі будемо вважати, що:

$$a_{r+1}^{k_r} = (a_r)_{k_r} \text{ та } f_R(x_h) = q(a_{r+1}^{k_r}), \forall k_r,$$

$$a_{r+2}^{k_{r+1}} = (a_{r+1})_{k_{r+1}} \text{ та } f_R(x_h) = q(a_{r+2}^{k_{r+1}}), \forall k_{r+1},$$

$$\dots \dots \dots$$

$$a_t^{k_{t-1}} = (a_{t-1})_{k_{t-1}} \text{ та } f_R(x_h) = q(a_t^{k_{t-1}}), \forall k_{t-1},$$

$$a_{t+1}^{k_t} = (a_t)_{k_t} \text{ та } f_R(x_h) = q(a_{t+1}^{k_t}), \forall k_t,$$

$$\dots \dots \dots$$

$$a_n^{k_{n-1}} = (a_{n-1})_{k_{n-1}} \text{ та } f_R(x_h) = q(a_n^{k_{n-1}}), \forall k_{n-1},$$

$$a_{n+1}^{k_n} = (a_n)_{k_n} \text{ та } f_R(x_h) = q(a_{n+1}^{k_n}), \forall k_n.$$

Зауважимо, що тут  $r \leq t \leq n - 1$ ,  $k_1, k_{r+1}, \dots, k_t, k_{t+1}, \dots, k_{n-1}, k_n$  – множина значень ознак  $P_{i_r}, P_{i_{r+1}}, \dots, P_{i_t}, P_{i_{t+1}}, \dots, P_{i_{n-1}}, P_{i_n}$  відповідно.

Крок N. Усі інші знайдені помилки розпізнавання у вибірці  $((x_{m+2}, f_R(x_{m+2})), \dots, (x_{m+\delta}, f_R(x_{m+\delta})))$  усуваються аналогічно.

**Твердження 2.1** Якщо  $\max_{0 \leq j \leq k-1} M_{\eta_{i_1} \dots \eta_{i_\xi}}^j = S$ , ( $1 \leq \xi \leq n$ ), то на шляху  $\eta_{i_1} \dots \eta_{i_\xi}$  ЛДК у процесі розпізнавання можливе не більше ніж  $K_{n-\xi} - S$  помилок, де  $K_{n-\xi} = K_1 * K_2 * \dots * K_n / K_{i_1} * \dots * K_{i_\xi}$ . Кількість схем ЛДК визначається рекурсивним відношенням:  $N_{n+1} = N_n^2 * (n + 1)$ , де  $N_2 = 2$ ,  $n$  – кількість ознак (ознак об'єкта в НВ) задачі розпізнавання.

**Теорема 2.1** Набір ознак із номерами  $i_1, \dots, i_\xi$ , ( $\xi \in 1, \dots, n$ ) є тестом тоді і тільки тоді, коли для них виконується:

$$\sum_{r_{i_1} \dots r_{i_\xi}} \left( \frac{M_{r_{i_1} \dots r_{i_\xi}}}{m} \right) * a_{r_{i_1} \dots r_{i_\xi}} = 1.$$

Ця формула дає можливість знаходити тести заданої довжини (під довжиною тесту розуміємо кількість ознак, що входять до нього).

**Зауваження.** Якщо хоча би один член даного виразу не дорівнює одиниці, то набір ознак з номерами  $r_{i_1} \dots r_{i_\xi}$  не є тестом.

У розділі досліджено питання побудови ЛДК за мінімальним тестом та подано відповідь на питання: чи буде таке побудоване дерево мінімальним у плані складності. ЛДК, побудоване за даними НВ, називається мінімальним, якщо не існує іншого ЛДК\*, побудованого за цією же НВ, у котрого кількість вершин менша за початкове дерево. Якщо група (набір) ознак з номерами  $r_{i_1} \dots r_{i_\xi}$  є мінімальним тестом,

то побудова ЛДК за даними ознаками не є важкою. Встановлено, що ЛДК, побудоване за мінімальним тестом, не завжди є мінімально можливим (Табл. 1).

Таблиця 1 – Залежність складності логічного дерева від кількості ознак

$KL \setminus D$	1	2	3	4	5	6	7	8	...	$n$
<i>Max</i>	3	7	15	31	63	127	255	511	...	$2^{n+1} - 1$
<i>Min</i>	3	5	7	9	11	13	15	17	...	$2n + 1$

Тут  $D$  – довжина тупикового тесту,  $KL$  – кількість можливих вершин,  $Max$  – вказує на максимальну кількість вершин ЛДК (складність), при заданій довжині тупикового тесту,  $Min$  – вказує на мінімальну кількість вершин ЛДК, можливу при заданій довжині тупикового тесту. Можливий випадок, коли ЛДК, побудоване за мінімальним тестом, може мати більше вершин, ніж ЛДК, побудоване за тупиковим тестом більшої довжини. Наприклад, ЛДК, побудоване за мінімальним тестом довжини 4, може мати 31 вершину, а ЛДК, побудоване за тупиковим тестом довжини 14, може мати 29 вершин, ЛДК, побудоване за мінімальним тестом довжини 10, може мати 2047 вершин, а ЛДК, побудоване за тупиковим тестом довжини 100, може мати 201 вершину (за тупиковим тестом довжиною 1000 – 2001 вершину).

Складність ЛДК залежить від кількості кінцевих вершин у дереві. Якщо кількість кінцевих вершин у дереві дорівнює  $K$ , то загальна кількість вершин  $N$  в структурі ЛДК можна розрахувати за формулою  $N = 2K - 1$ . Це справедливо у разі, коли  $(P_i(x) \in \{0,1\})$ ,  $2 \leq K \leq 2^n$ , де  $n$  – кількість ознак об'єктів НВ.

Отже, в розділі розроблено схему виправлення помилок ЛДК (схема УПР) 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 НВ.

У **третьому розділі** введено  $T$  – опорні множини, які представляють собою відібраний набір ознак з фіксованими їх значеннями:  $(\begin{matrix} e_{i_1} & e_{i_k} \\ X_{i_1} & \dots & X_{i_k} \end{matrix})$ ,  $k = 1, \dots, n$  на основі інформації деякої початкової НВ, з можливістю наступної оцінки даних опорних множин за допомогою відповідних функціоналів. Уводяться системи  $T$  – опорних множин різних типів.

Запропоновано методи задання систем  $T$  – опорних множин:

1) Базовий спосіб задання систем  $T$  – опорних множин  $\Omega^T = \{w_1, \dots, w_\xi\}$  – полягає у використанні їх основного визначення. Випадковим чином фіксуються піднабори ознак та їх значення, а, використовуючи програмний генератор псевдо-випадкових чисел, можна сформулювати різні системи  $T$  – опорних множин. На наступному етапі проводиться процедура селекції та виділення різних типів систем  $T$  – опорних множин  $\Omega^T = \{w_1, \dots, w_\xi\}$ .

2) Задання систем  $T$  – опорних множин  $\Omega^T = \{w_1, \dots, w_\xi\}$  за допомогою наборів тестів. Нехай ознаки з номерами  $i_1, \dots, i_k$  утворюють тест, знайдений за деякою початковою інформацією  $I(l)$ . Якщо зафіксувати значення ознак, що входять

до тесту, то дана множина буде утворювати несуперечливу та не завжди інформативну систему  $T$  – опорних множин  $\Omega^T = \{w_1, \dots, w_\xi\}$ .

3) Є можливість знаходити системи  $T$  – опорних множин  $\Omega^T = \{w_1, \dots, w_\xi\}$ , у тому числі інформативні, несуперечливі, прості, використовуючи інші методи та алгоритми (метод випадкової генерації).

4) Використовуючи поняття допустимого розбиття та схему його загальної побудови, можна знаходити системи несуперечливих  $T$  – опорних множин  $\Omega^T = \{w_1, \dots, w_\xi\}$  за інформацією  $I(l)$  з числом  $T$  – опорних множин не більше деякого  $l$ , де  $l$  – кількість об'єктів в початковій інформації  $I(l)$ .

5) Знаходження простих систем  $T$  – опорних множин  $\Omega^T = \{w_1, \dots, w_\xi\}$  за допомогою  $I(l)$  та дерев розпізнавання (структур ЛДК) має найбільше перспектив, оскільки мало залежить від об'єму та виду початкової інформації  $I(l)$ , тобто ЛДК є стисненим структурним типом даних.

Подано визначення алгоритму розпізнавання на основі  $T$  – опорної множини:

$$\{A\} = (\Omega_A^T, B(w, S), \Gamma_j(S), y = f(\Gamma_j(S))).$$

Де  $\Omega_A^T$  – деяка фіксована система  $T$  – опорних множин  $\Omega^T = \{w_1, \dots, w_\xi\}$ , яка побудована за початковою інформацією  $I(l)$ ,  $S(a_1, \dots, a_n)$  – деякий об'єкт з множини об'єктів  $M$ ,  $B(w, S)$  – функція близькості, інформативності об'єкта  $S$  до  $T$  – опорної множини  $w$ . На наступному етапі за  $\Omega_A^T$  та об'єктом  $S$ , будуються оцінки  $\Gamma_j(S)$  належності  $j$  – вому класу об'єкту  $S$  наступним чином:  $\Gamma_j(S) = \sum_{w \in \Omega_A^T} B(w, S)$ . Класифікація об'єкта  $S$  проводиться за допомогою деякої скінчено-значної функції  $y = f(\Gamma_j(S))$ , що приймає значення  $0, 1, \dots, l$ .

У розділі досліджено основні підмоделі алгоритму розпізнавання на основі  $T$  – опорної множини.

1) Якщо в якості  $\Omega_A^T$  вибрати тільки інформативні несуперечливі системи  $T$  – опорних множин, то отримаємо системи  $T$  – опорних множин  $\Omega_A^T(U, H)$  та модель:  $R(U, H) = (\Omega_A^T(U, H), B(w, S), \Gamma_j(S), y = f(\Gamma_j(S)))$ .

**Теорема 3.1** Не існує деякого алгоритму  $A \in R(U, H)$ , який би видавав помилки на даних початкової навчальної інформації  $I(l)$  та відмови класифікації на всій множині допустимих об'єктів.

2) Якщо в якості  $\Omega_A^T$  вибрати тільки інформативні  $T$  – опорні множини, то отримаємо системи  $T$  – опорних множин  $\Omega_A^T(U, -)$  та модель:  $R(U, -) = (\Omega_A^T(U, -), B(w, S), \Gamma_j(S), y = f(\Gamma_j(S)))$ . Алгоритми цієї моделі можуть видавати помилки розпізнавання на даних початкової навчальної інформації  $I(l)$  та відмови класифікації на деяких допустимих об'єктах.

3) Якщо в якості  $\Omega_A^T$  вибрати тільки несуперечливі  $T$  – опорні множини, то отримаємо системи  $T$  – опорних множин  $\Omega_A^T(-, H)$  та модель:  $R(-, H) = (\Omega_A^T(-, H), B(w, S), \Gamma_j(S), y = f(\Gamma_j(S)))$ . Алгоритми цієї моделі не дають помилок розпізнавання на даних початкової навчальної інформації  $I(l)$ , проте дають відмови класифікації на деяких допустимих об'єктах.

4) Якщо в якості  $\Omega_A^T$  вибрати просто довільні  $T$  – опорні множини, то отримаємо системи  $T$  – опорних множин  $\Omega_A^T(-, -)$  та модель:  $R(-, -) = (\Omega_A^T(-, -), B(w, S), \Gamma_j(S), y = f(\Gamma_j(S)))$ .

**Теорема 3.2** Кількість усіх алгоритмів у моделі  $R(-, -)$  дорівнює  $2^{3^n-1}$ , де  $n$  – кількість ознак у задачі розпізнавання.

Встановлено, що ємність усіх моделей  $R(-, -)$ ,  $R(-, H)$ ,  $R(U, -)$ ,  $R(U, H)$  є скінченною.

У розділі досліджено питання взаємозв'язку  $T$  – опорних множин з структурами ЛДК. Правилком розпізнавання являється деякий оператор  $\Pi$ , що приводить довільний допустимий об'єкт  $S$  за його інформаційним описом  $I(S)$  в кінцеву вершину ЛДК. Деякий алгоритм  $A$  називається алгоритмом розпізнавання (класифікації), якщо він переводить початкову інформацію  $I(l)$  та інформаційний опис довільного числа  $q$  допустимих об'єктів  $I(q)$  у вектор  $a_{r+q}$ , складений з елементів  $\{0, 1, \dots, l\}$  де  $r = |I(l)|$  – кількість об'єктів, що складають початкову інформацію  $I(l)$ , при чому об'єктам із  $I(l)$  відповідає значення функції, що задає розбиття, при умові початкової відмінності між об'єктами різних класів. Фіксоване ЛДК, побудоване за початковою інформацією  $I(l)$  із заданим правилом розпізнавання, задає деякий алгоритм розпізнавання  $A$ .

**Теорема 3.3** Довільне ЛДК задає систему  $T$  – опорних множин, причому кількість  $T$  – опорних множин у цій системі дорівнює кількості кінцевих вершин ЛДК, а кількість інформативних  $T$  – опорних множин до заданих класів дорівнює кількості визначених вершин ЛДК – (Рис. 2). Для наведеного прикладу загальна кількість  $T$  – опорних множин складає 8, причому серед них будуть інформативні до класу  $H_1$  – 2, інформативні до класу  $H_2$  – 3, інформативні до класу  $H_3$  – 3.

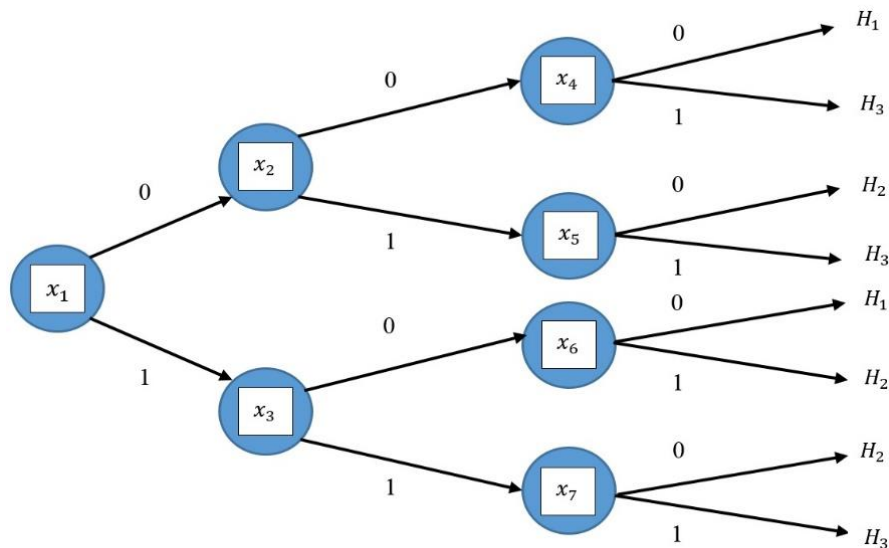


Рисунок 2 – Структура ЛДК, яке задає вісім  $T$  – опорних множин.

У роботі введено дерева моделей класифікації, які представляють собою зв'язаний граф без циклів, в некінцевих вершинах якого знаходяться фіксовані моделі з  $\{I_j^{(nk)}, P_{nk}\}$ , а ребра нумеруються значеннями предикатів з цих моделей. У кінцевих вершинах розташовані значення функції, що задає розбиття  $I(l)$  на класи або символ  $\Delta$  невизначеності. Причому на фіксованому шляху такої конструкції одна й та ж сама модель може зустрічатися тільки один раз. Довільне ДМК являється алгоритмом класифікації (у вигляді логічного дерева) деяких об'єктів за початковою інформацією  $IV$ .

**Визначення 3.1** Деяка модель  $\{I(j^{\eta_k}), P_{\eta_k}\}$  називається інформативною по відношенню до об'єкту  $S$ , якщо значення предиката  $P_{\eta_k}$  вірно на його частковому (ознаковому) описі.

**Визначення 3.2** Модель  $\{I(j^{\eta_k}), P_{\eta_k}\}$  називається інформативною для деякого класу  $j$  за початковою інформацією  $I(l)$ , якщо вона інформативна тільки для деяких об'єктів цього класу. Відносно об'єктів, що не входять в  $I(l)$ , ніяких обмежень не накладається.

У розділі введено схеми обчислення міри, що характеризує роздільні можливості окремих ознак (інформативність) відносно  $f_R(s_i)$  за початковою інформацією  $R(I(l))$ :

$$W(P_i) = \frac{1}{M} \sum_{j=G_i} \max_{0 \leq k \leq l} b_j^k.$$

Тут  $M$  – множина об'єктів  $I(l)$ ,  $G_i$  – множина значень  $P_i$  – тової ознаки,  $b_j^k$  – кількість значень  $j$  -  $P_i$  – тової ознаки в класі  $H_k$ , ( $k = 1, \dots, l$ ),  $l$  – кількість класів у  $H$ .

$$W(P_i \setminus P_{i_1} = \eta_1, \dots, P_{i_t} = \eta_t) = \frac{1}{h} \sum_{j=G_i} \max_{0 \leq k \leq l} b_j^k.$$

Тут  $i \neq i_1, \dots, i_t$ , беруться тільки ті об'єкти, для яких  $P_{i_1} = \eta_1, \dots, P_{i_t} = \eta_t$ ,  $h$  – кількість таких об'єктів.

$$W(P_{i_1}, \dots, P_{i_t}) = \frac{1}{M} \sum_{\Delta_j=G_{\Delta_j}} \max_{0 \leq k \leq l} b_{\Delta_j}^k.$$

Тут  $\Delta_j = (x_{i_1}, \dots, x_{i_t})$ , ( $x_{i_k} \in G, k = 1, \dots, t$ ) – фіксований набір значень ознак  $P_{i_1}, \dots, P_{i_t}$ ,  $G_{\Delta_j}$  – множина значень ознак, що відрізняються між собою значеннями хоча би однієї ознаки,  $b_{\Delta_j}^k$  – кількість  $\Delta_j$  в класі  $H_k$ , ( $k = 1, \dots, l$ ).

$$W(P_{i_1}, \dots, P_{i_t} | P_{i_1} = \eta_1, \dots, P_{i_r} = \eta_r) = \frac{1}{h} \sum_{\Delta_j=G_{\Delta_j}} \max_{0 \leq k \leq l} b_{\Delta_j}^k.$$

Тут беруться тільки ті об'єкти, для яких  $P_{i_1} = \eta_1, \dots, P_{i_r} = \eta_r$ ,  $h$  – кількість таких об'єктів.

Отже, у розділі введено дерева моделей для задачі класифікації. Показано, що деяка множина моделей  $\{I(j^{\eta_k}), P_{\eta_k}\}$ , що представляють алгебраїчні системи розпізнавання (у вигляді структур дерев класифікації) при фіксованому  $\eta_k$  та  $P_{\eta_k}$ , перетворюється на конкретну модель класифікації  $\{I(j^{\eta_k}), P_{\eta_k}\}$  (деревоподібну логічну структуру), причому правилом розпізнавання (класифікації) дерева моделей являється деякий оператор, що відносить довільний допустимий об'єкт  $S$  (за його початковим інформаційним описом) у фіксовану кінцеву вершину дерева моделей. Запропоновано використання концепції  $T$  – опорних множин для опису дискретних об'єктів, на основі чого представлено нові означення інформативності по відношенню до класу та об'єкту класифікації, причому важливим моментом є те, що  $2T$  – опорні множини можна використовувати в якості ознак, що характеризують деякий об'єкт, а відповідно для побудови алгоритмів розпізнавання або класифікації дискретних об'єктів доречно використовувати дерева моделей класифікації (алгебраїчних систем), у вершинах яких знаходяться  $2T$  – опорні множини. Запропоновано якісні схеми для обчислення деякої роздільної можливості (інформативності) окремих дискретних ознак відносно ФР за початковою інформацією  $I(l)$ , причому цей набір функціоналів оцінки якості дискретних ознак дозволяє побудувати мінімальне за ознаковим описом представлення дискретного об'єкту довільної структури.



Підкреслено, що питання мінімізації опису дискретних об'єктів тісно пов'язано з функціональною оцінкою важливості дискретних ознак (які характеризують даний об'єкт), що дозволяє максимально компактно та якісно описати об'єкт класифікації.

У четвертому розділі показано, для  $k$  – значної функції  $f(x_1, \dots, x_n)$  від  $n$  змінних, заданої таблично та визначеної на всіх наборах, можна побудувати відповідне повне логічне дерево, мінімізувати його і записати аналітичний вигляд, а за табличним представленням знайти аналітичний вигляд цієї функції у вигляді ДНФ або КНФ та побудувати логічне дерево (Рис. 3). Показано, що за підмножиною будь-яких значень набору змінних можна побудувати повне логічне дерево, яке представляє функцію розпізнавання, визначену на всіх наборах, це дозволяє застосовувати логічні дерева в розпізнаванні образів, причому одержання тієї чи іншої логічної функції залежить від порядку проходження змінних у логічному дереві, що безпосередньо впливає на його кінцеву структуру.

**Теорема 4.1** Довільну логічну функцію  $f(x_1, x_2, \dots, x_n)$   $k$  – значної логіки від  $n$  змінних можна представити у вигляді:

$$f(x_1, x_2, \dots, x_n) = \varphi_0(x_i)f(x_1, x_2, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n) \vee \\ \vee \varphi_1(x_i)f(x_1, x_2, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_n) \vee \dots \vee \\ \vee \varphi_{k-1}(x_i)f(x_1, x_2, \dots, x_{i-1}, k-1, x_{i+1}, \dots, x_n)$$

Зауважимо, тут  $f \in \{0,1, \dots, k-1\}, x_i \in \{0,1, \dots, k-1\}, (i = 1,2, \dots, n)$ .

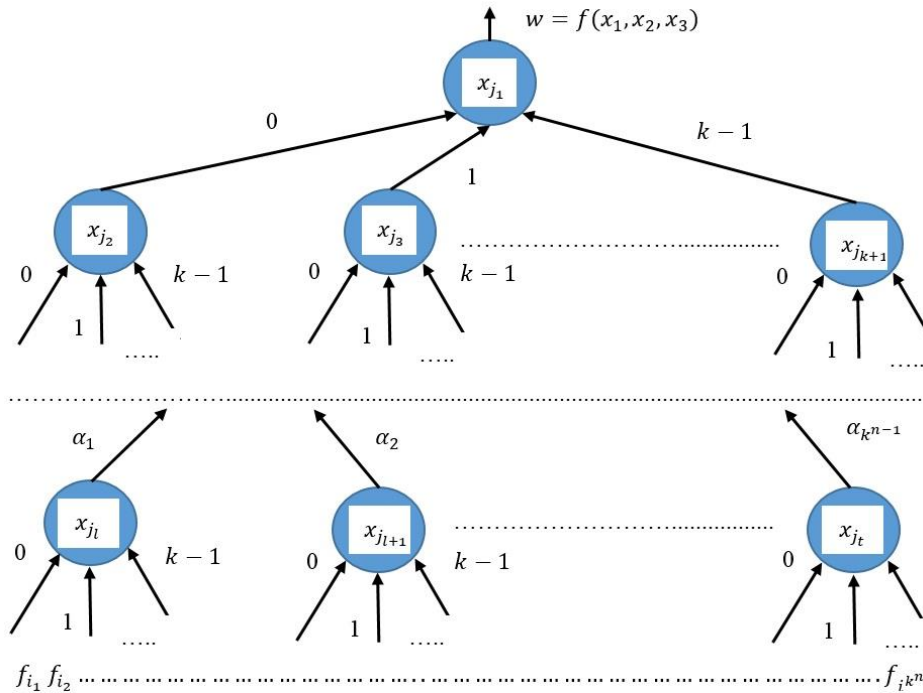


Рисунок 3 – Логічне дерево (граф) функції  $k$  – значної логіки від  $n$  змінних.

**Визначення 4.1** Нехай задано довільну функцію двох змінних  $f(x_1, x_2)$  багатозначної логіки. Трикутником будемо називати логічне дерево, яке можна побудувати для цієї функції (Рис. 4).

Базова теорема в схемі перестановки ярусів структури логічного дерева має наступний вигляд:

**Теорема 4.2**

$$x_1 * (x_2 * (\alpha_0^0, \alpha_1^0, \dots, \alpha_{k-1}^0); x_2 * (\alpha_0^1, \alpha_1^1, \dots, \alpha_{k-1}^1); \dots \\ \dots; x_2 * (\alpha_0^{k-1}, \alpha_1^{k-1}, \dots, \alpha_{k-1}^{k-1})) = x_2 * (x_1 * (\alpha_0^0, \alpha_1^0, \dots, \alpha_{k-1}^0); \\ ; x_1 * (\alpha_1^0, \alpha_1^1, \dots, \alpha_1^{k-1}); \dots; x_1 * (\alpha_{k-1}^0, \alpha_{k-1}^1, \dots, \alpha_{k-1}^{k-1})).$$

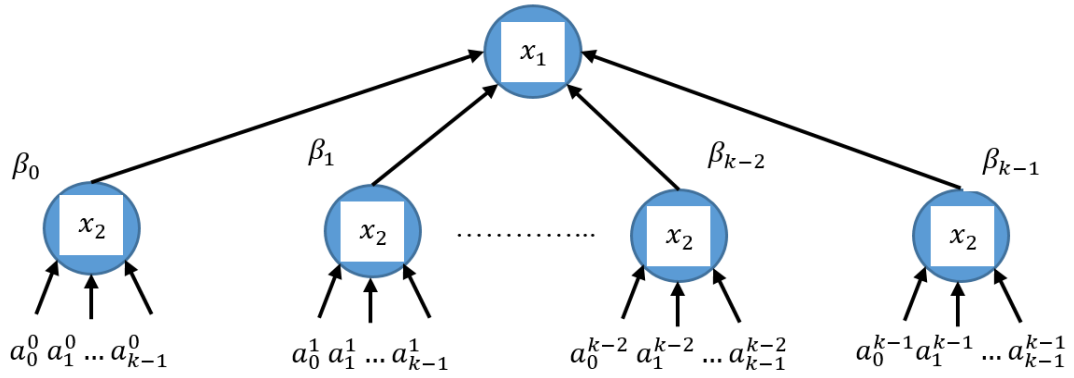


Рисунок 4 – Трикутник (логічне дерево) функції  $f(x_1, x_2)$ .

Теорема визначає наступний факт – що при перестановці ярусів логічного дерева, відповідних змінних  $x_1$  та  $x_2$ , вхідні значення  $\alpha_0^0, \dots, \alpha_{k-1}^0, \alpha_0^1, \alpha_1^1, \dots, \alpha_{k-1}^1, \alpha_{k-1}^1, \dots, \alpha_{k-1}^{k-1}$  міняються місцями та представляються у вигляді  $\alpha_0^0, \dots, \alpha_0^{k-1}, \alpha_1^0, \dots, \alpha_{k-1}^0, \dots, \alpha_{k-1}^{k-1}$  – (Рис. 5).

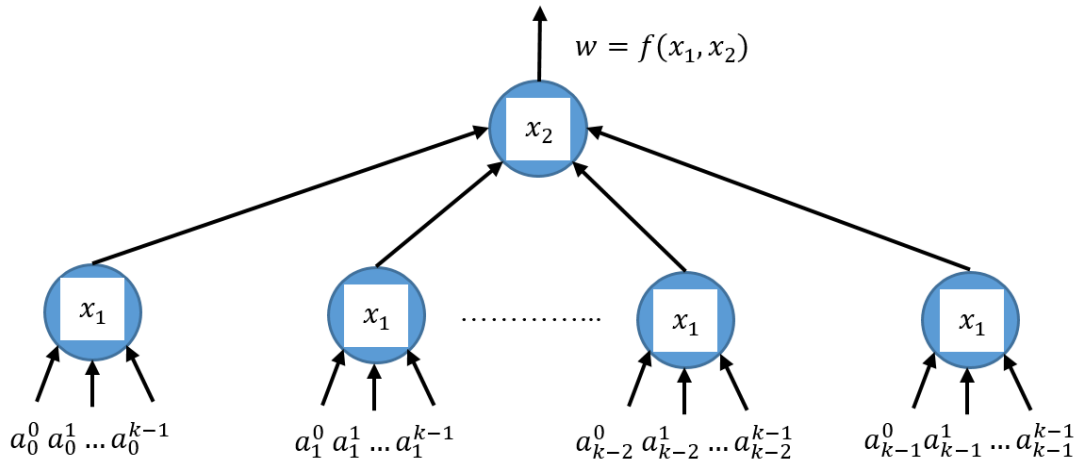


Рисунок 5 – Трикутник (логічне дерево) функції  $f(x_1, x_2)$  при перестановці ярусів, відповідних змінних  $x_1$  та  $x_2$ .

Зафіксовано, що мінімізація отримується за рахунок зменшення кількості різних функцій (міток, атрибутів) у структурі логічного дерева.

Схема мінімізації (обрізки) логічного дерева полягає у наступному: на вході маємо побудовану структуру логічного дерева. Розглядаємо трикутник, який утворюється першим та другим каскадами, використовуючи властивості трикутника, якщо потрібно переставляємо змінні  $x_1$  та  $x_2$ . На наступному етапі перевіряємо трикутники, які утворені другим та третім каскадами іншими змінними. Застосуємо аналогічні судження до кожного трикутника, спускаючись вниз за ярусами, доки не перевіriamo всі трикутники.

Схема мінімізації логічної функції на основі логічного дерева полягає у наступному:

- 1) За табличною (або іншою) формою задання довільної логічної функції  $n$  змінних  $f(x_1, x_2, \dots, x_n)$  побудувати відповідне логічне дерево.
- 2) На отриманій структурі логічного дерева провести розстановку відповідних міток (атрибутів) дерева та проаналізувати можливий ефект (щодо оптимізації структури дерева) від перестановки в його структурі ярусів.

3) Використовуючи теорему 4.1 (про розклад функції  $n$  змінних  $f(x_1, x_2, \dots, x_n)$  за фіксованою змінною), необхідно представити логічну функції через відповідні мітки (атрибути).

4) Усі мітки логічного дерева відповідної функції  $n$  змінних  $f(x_1, x_2, \dots, x_n)$  необхідно представити через змінні  $x_1, x_2, \dots, x_n$ , тобто отримаємо деяке дужкове представлення логічної функції.

5) В отриманому представленні логічної функції  $n$  змінних  $f(x_1, x_2, \dots, x_n)$  на основі базових співвідношень проводиться процес спрощення та інші дужкові перетворення.

У розділі представлено конструкцію найскладнішого логічного дерева (Рис. 6).

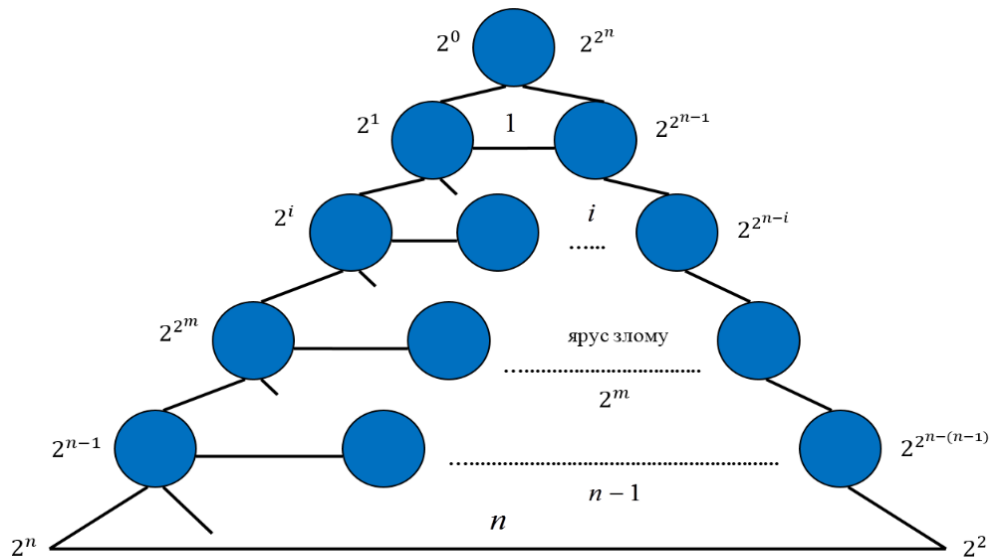


Рисунок 6 – Загальна структура найскладнішого логічного дерева.

Під ярусом злому розуміємо такий ярус деякого логічного дерева за номером  $i$ , де всі функції різні в усіх вершинах (вузлах) дерева, тобто -  $2^i = 2^{2^{n-i}}$ .

У розділі досліджено питання впливу перестановки ярусів на структурну складність логічного дерева. Так, на першому етапі перестановки ( $2^m$ ) – ого та ( $2^m - 1$ ) – ого ярусу структура логічного дерева буде наступною (Рис. 7):

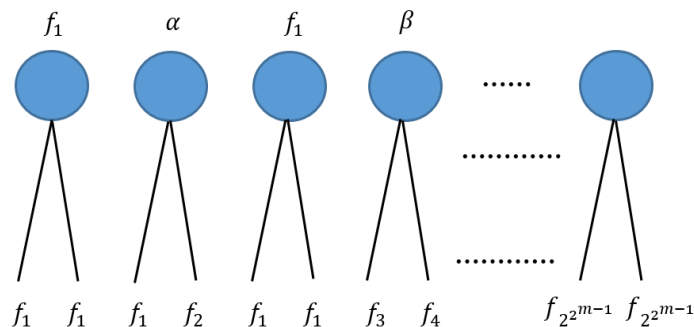


Рисунок 7 – Початковий етап схеми перестановки ярусів у логічному дереві.

У цьому разі на ярусі ( $2^m - 1$ ) отримаємо таке саме розташування міток, яке було на ( $2^m$ ) ярусі. На наступних етапах проводиться перестановка ярусів ( $2^m - 1$ ) – й та ( $2^m - 2$ ) – й і так далі, до ( $2^m - 1$ ) – го ярусу. Схемою перестановки ярусів на виході отримаємо структуру, починаючи з ( $2^{m-1} - 1$ ) ярусу логічного дерева вигляду, поданого на Рис. 8.

Загальна кількість усіх міток після проведення всіх перестановок буде дорівнювати:  $2^{2^{m-1}} - 1 + 2^{2^{m-1}+1} - 1 - 1 - 2^{2^{m-1}} - 1 = 3 * 2^{2^{m-1}} - 1$ .

При  $m = 2$  до проведення перестановок ярусів маємо  $2^{2^2+1} - 1 = 31$  мітку, а після наведеної процедури:  $3 * 2^{2^{2-1}} - 1 = 11$ .

Зафіксовано, що після проведення процедури перестановок маємо виграш у складності фінальної структури логічного дерева майже в три рази.

Дано числову оцінку виграшу в складності структури логічного дерева для загального випадку:

$$\frac{2^{2^{m+1}}}{3 * 2^{2^{m-1}}} = \frac{1}{3} * 2^{2^{m-1}+1}.$$

При  $m = 3$  маємо  $\frac{1}{3} * 2^{2^{3-1}+1} \cong 10$ , при  $m = 4$  маємо  $\frac{1}{3} * 2^{2^{4-1}+1} \cong 160$ .

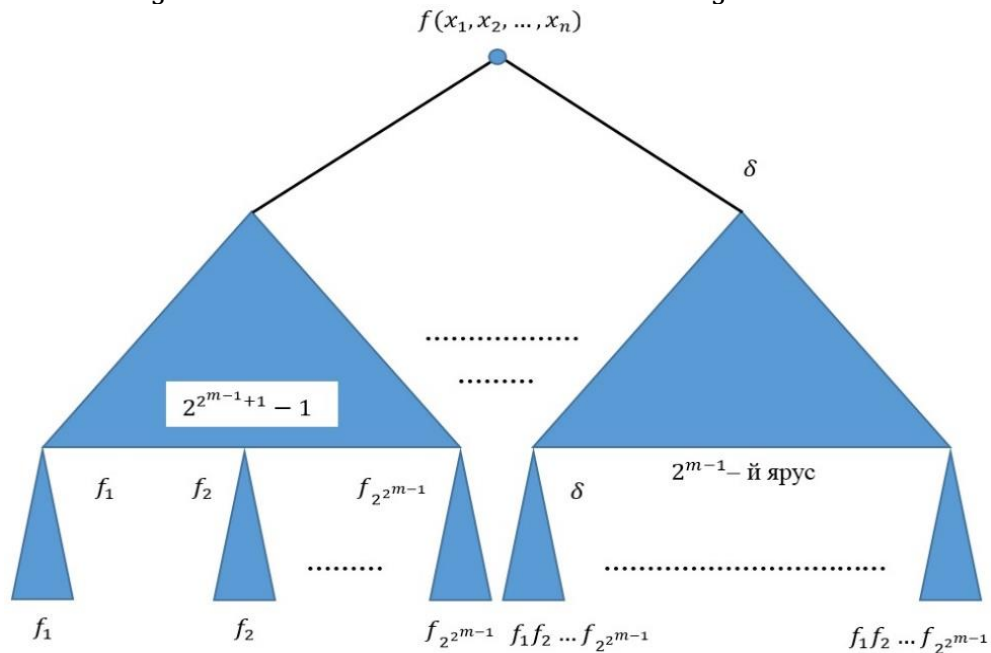


Рисунок 8 – Структура ЛДК після проведення процедури перестановки ярусів.

Зі зростанням числа  $m$ , та відповідно  $n$ , ефективність схеми перестановки ярусів логічного дерева швидко зростає. Наприклад, при  $m = 5$  маємо наступне:  $\frac{1}{3} * 2^{2^{5-1}+1} = \frac{2^{17}}{3}$ , при цьому число  $n = 2^5 + 5 = 37$ .

Отже, в розділі запропоновано простий метод мінімізації дискретних структур на основі концепції логічних дерев (процедури перестановки ярусів у його структурі), причому даний підхід характеризується простотою роботи за вказаною схемою на відповідній графічній структурі (конструкції логічного дерева), значною наочністю порівняно з іншими методами (є прямим або опосередкованим узагальненням двійкових (бінарних) методик для багатозначного випадку), можливістю нескладної програмної реалізації, що значно розширює сферу дій для великих  $n$  та  $k$  у випадку багатозначної логіки. Наведено числову оцінку впливу процедури перестановки ярусів структури регулярного логічного дерева на його складність для бінарного випадку, причому ефект перестановки ярусів є значним. Запропонований важливий механізм мінімізації логічних дерев – процедура перестановки ярусів (блоків) у структурі дерева – дає змогу досягти помітного ефекту зменшення складності логічного дерева.

У п'ятому розділі побудовано схему найскладнішого логічного дерева, на основі критерію структурної складності, яке містить у своїй конструкції максимальну кількість різних міток (атрибутів, вершин, функцій). Для довільного регулярного дерева показано, що якщо на фіксованому ярусі в усіх вершинах стоять різні функції, мітки, тоді в усіх вершинах ярусів, які розташовані вище в структурі цього логічного дерева також стоять різні функції. Представлено загальну числову оцінку структурної складності, кількості різних міток, функцій, атрибутів найскладнішого логічного дерева залежно від розташування ярусу злomu в його структурі.

**Теорема 5.1** Якщо на  $l$ -му ярусі деякого логічного дерева  $D$  в усіх вершинах стоять різні функції, мітки, атрибути, тоді в усіх вершинах  $0, 1, 2, \dots, l-1$  ярусів даного логічного дерева стоять різні функції.

Схема розрахунку величини структурної складності  $S_{n,k}$  логічного дерева  $D_{n,k}^*$  полягає в наступному:

1) Для випадку структури  $k^{l_0} = k^{k^{n-l_0}}$  маємо  $n = k^m$ .

Отже  $n - m = n - \log_k n$ .

Величина  $S_{n,k}$  – буде представлятися наступною формулою:

$$S_{n,k} = 1 + k + k^2 + \dots + k^{l_0} = \frac{k^{l_0+1} - 1}{k - 1}.$$

Для представлення  $l_0 = n - \log_k n$  маємо:  $S_{n,k} = \frac{k^{n+1}}{n*(k-1)} - \frac{1}{k-1}$ .

Структурна складності  $S_{n,k}$  логічного дерева  $D_{n,k}^*$  у цьому разі:

$$S_{n,k} \approx \frac{k^{n+1}}{n * (k - 1)}.$$

2) Для випадку структури  $k^{l_0} \neq k^{k^{n-l_0}}$ .

Маємо, що  $1 + k + \dots + k^{l_0} \leq S_{n,k} \leq 1 + k + \dots + k^{l_0} + 1$ , та відповідно  $\frac{k^{l_0+1}-1}{k-1} \leq S_{n,k} \leq \frac{k^{l_0+2}-1}{k-1}$ . Тоді  $S_{n,k} = \frac{k^{l_0+1+r}-1}{k-1}$ , де величина  $0 \leq r \leq 1$ .

Враховуючи, що  $l_0 = n - \log_k n + j$ , де  $-1 \leq j < 1$ , величина структурної складності  $S_{n,k} = \frac{k^{1+n-\log_k n+\rho}-1}{k-1}$ , де  $-1 \leq \rho < 2$ .

Величина  $S_{n,k}$  – буде представлятися наступною формулою:

$$S_{n,k} = \frac{\frac{k^{1+n+\rho}-1}{n}-1}{k-1} = \frac{k^{1+n+\rho}}{n*(k-1)} - \frac{1}{k-1}, \text{ де } -1 \leq \rho < 2.$$

Структурна складності  $S_{n,k}$  логічного дерева  $D_{n,k}^*$  для даного випадку:

$$S_{n,k} \approx \frac{k^{1+n+\rho}}{n*(k-1)}.$$

Для всіх випадків у даній схемі число  $S_{n,k}$  – ціле.

У розділі розроблено методи знаходження подібності  $S_n$  для конструкцій логічних дерев у задачах мінімізації їх структур, та на основі цього досліджено питання критерію оптимальності регулярного логічного дерева. Розв'язок даного питання має принципову важливість з точки зору як пошуку ефективних перестановок в процедурі мінімізації структур регулярних логічних дерев, так і оцінки загального ефекту результату такої оптимізації логічних дерев.

Базова схема знаходження подібності  $S_n$  для конструкції логічного дерева полягає в наступному. Задано логічне дерево функції  $f(x_1, x_2, \dots, x_n)$ . При цьому можливі два випадки:

а) Верхня вершина (вершина першого ярусу) логічного дерева є подібною. В цьому разі логічне дерево можна представити піддеревом – лівою гілкою початкового дерева. Нехай верхня вершина логічного дерева відповідає змінній  $x_i$ . Тоді має місце наступна формула:

$$L_{\text{дуж}}(f) = L_{\text{дуж}}(f_{\varphi_0}(x_i)) = \frac{k^n - k}{k - 1} - S_{n-1}^{\varphi_0^i},$$

Тут  $S_{n-1}^{\varphi_0^i}$  – подібність піддерева, яке представляє ліву гілку (піддереву) початкового логічного дерева. В цьому випадку будемо мати наступне:

$$S_n = k^n + S_{n-1}^{\varphi_0^i}.$$

б) Верхня вершина логічного дерева (вершина першого ярусу) не є подібною, тоді будемо мати наступне:

$$L_{\text{дуж}}(f) = L_{\text{дуж}}(f_{\varphi_0}(x_i)) + L_{\text{дуж}}(f_{\varphi_1}(x_i)) + L_{\text{дуж}}(f_{\varphi_{k-1}}(x_i)) + k.$$

Кількість міток кожного з піддерев початкового логічного дерева визначається наступним чином:

$$L_{\text{дуж}}(f_{\varphi_0}(x_i)) = \frac{k^n - k}{k - 1} - S_{n-1}^{\varphi_0^i},$$

$$L_{\text{дуж}}(f_{\varphi_1}(x_i)) = \frac{k^n - k}{k - 1} - S_{n-1}^{\varphi_1^i},$$

.....

$$L_{\text{дуж}}(f_{\varphi_{k-1}}(x_i)) = \frac{k^n - k}{k - 1} - S_{n-1}^{\varphi_{k-1}^i}.$$

Тут  $S_{n-1}^{\varphi_0^i}, S_{n-1}^{\varphi_1^i}, \dots, S_{n-1}^{\varphi_{k-1}^i}$  – подібність відповідних піддерев (підграфів) початкового логічного дерева. Отже, будемо мати наступне:

$$\begin{aligned} L_{\text{дуж}}(f) &= \frac{k^n - k}{k - 1} - S_{n-1}^{\varphi_0^i} + \frac{k^n - k}{k - 1} - S_{n-1}^{\varphi_1^i} + \dots + \frac{k^n - k}{k - 1} - S_{n-1}^{\varphi_{k-1}^i} + k = \\ &= \frac{k^n - k}{k - 1} * k + k - S_{n-1}^{f_i} = \frac{k^{n+1} - k^2 + k^2 - k}{k - 1} - S_{n-1}^{f_i} = \frac{k^{n+1} - k}{k - 1} - S_{n-1}^{f_i}. \end{aligned}$$

Тоді отримаємо  $S_{n-1}^{f_i} = \sum_{j=1}^{k-1} S_{n-1}^{\varphi_j^i}$ ,  $S_n = S_{n-1}^{f_i}$ .

Об'єднавши результати пунктів (а) та (б), отримаємо фінальну схему:

$$S_n = \begin{cases} k^n + S_{n-1}^{\varphi_0^i}, & \text{якщо верхня вершина дерева є подібною.} \\ S_{n-1}^{f_i}, & \text{якщо верхня вершина дерева не є подібною.} \end{cases}$$

У розділі запропоновано схеми оптимального розташування змінних в структурі логічного дерева, які в більшості випадків дають оптимальне логічне дерево (або близьке до оптимального відносно його структури). Показано, що оптимальне логічне дерево (деякої функції з фіктивною змінною) можна побудувати, записавши в його верхню вершину номер фіктивної змінної, тобто оптимальне дерево функції з фіктивною змінною можна завжди зробити особливим логічним деревом, що в

перспективі забезпечує можливість провести процедуру ефективної мінімізації його структури.

Схема (А) оптимального розташування міток структур логічних дерев (на основі подібності вершин структури ЛДК) полягає в наступному. До верхньої вершини логічного дерева записуємо змінну  $x_i$  з мінімальною величиною  $S_f^i$ . Якщо таких вершин декілька, то записується будь-яка з них. У вершинах піддерев функцій  $f_{\varphi_0}(x_i), f_{\varphi_1}(x_i), \dots, f_{\varphi_{k-1}}(x_i)$  записуються відповідно, номери змінних з мінімальними величинами  $S_{f_{\varphi_l}}^j(x_i)$ , де  $l \in \{0, 1, 2, \dots, k-1\}$ ,  $j \in \{1, 2, \dots, n\}$ . Відмітимо, що якщо мінімальне значення цих величин однаково для декількох змінних, то записується будь-яка з них. Аналогічним чином знаходяться номери змінних для всіх інших вершин даного логічного дерева. В результаті застосування даної схеми отримується нерегулярне логічне дерево (граф), де всі його особливі піддерева складаються з подібних вершин.

Схема (В) оптимального розташування міток структур логічних дерев (на основі подібності вершин структури ЛДК) полягає в наступному. Якщо функція  $f(x_1, x_2, \dots, x_n)$  має фіктивну змінну, то фіктивна змінна розміщується в верхній вершині відповідного логічного дерева (піддерева). Якщо функція фіктивних змінних не має, то в верхню вершину логічного дерева розміщуємо змінну з мінімальною величиною  $S_f^i$ . У вершини піддерев функцій  $f_{\varphi_0}(x_i), f_{\varphi_1}(x_i), \dots, f_{\varphi_{k-1}}(x_i)$  записуються номери змінних, фіктивних для даних функцій, а якщо таких немає, то записується змінна  $x_l$  з мінімальною величиною  $S_{f_{\varphi_l}}^j(x_i)$ , де відповідно  $l \in \{0, 1, 2, \dots, n\}$ ,  $j \in \{0, 1, 2, \dots, k-1\}$ . Якщо змінних, які задовольняють вказаним вище умовам декілька, то у відповідну вершину логічного дерева записується довільна з них. Аналогічним чином знаходяться змінні для всіх інших вершин. Тут змінна  $x_l$  є фіктивною для функції  $f_{r_i}$ , якщо  $S_{f_{r_i}}^l = k^{n-t-1}$ , де  $t$  – кількість характеристичних функцій в функції  $r_i$ . Використовуючи це зауваження, легко можна визначити, яку змінну помістити до вершини дерева.

**Теорема 5.2** Якщо логічна функція  $f(x_1, x_2, \dots, x_n)$  має тільки одну фіктивну змінну  $x_1$ , для довільного неособливого логічного дерева можна побудувати особливе дерево з тою самою подібністю, в верхній вершині якого розташована змінна  $x_1$ .

Зауважимо, що виходячи з даної теореми, метод (В) буде давати кращій результат щодо побудованої структури логічного дерева ніж метод (А).

У шостому розділі запропоновано концепцію покрокової апроксимації масиву початкових даних НВ набором відібраних та оцінених незалежних алгоритмів класифікації та розпізнавання, яка дає можливість будувати різноманітні моделі АДК. Досліджено питання оцінки якості (ефективності, інформативності) набору алгоритмів класифікації (вершин структури дерева класифікації) в схемі методів АДК – відбору критерію розгалуження конструкції дерева класифікації.

У розділі вводяться базові критерії розгалуження для дерев алгоритмів (моделей АДК) наступного вигляду:

$$\delta_{a_1, \dots, a_i} = \frac{S_{a_1, \dots, a_i}}{m}, \quad \psi_{a_1, \dots, a_i}^j = \frac{S_{a_1, \dots, a_i}^j}{S_{a_1, \dots, a_i}}, \quad \rho_{a_1, \dots, a_i} = \max_j \psi_{a_1, \dots, a_i}^j.$$

Тут величина  $\delta_{a_1, \dots, a_i}$  характеризує частоту входжень членів послідовності  $x_1, x_2, \dots, x_m$  (дискретних об'єктів) в множину  $G_{a_1, \dots, a_i}$ , величина  $\psi_{a_1, \dots, a_i}^j$  характеризує частоту приналежності деякого об'єкту  $x$  класу  $H_j$  при умові, що  $x \in G_{a_1, \dots, a_i}$ . Умова  $x \in G_{a_1, \dots, a_i}$  еквівалентна умові, що в послідовності алгоритмів  $a_1, \dots, a_i$  знайдеться такий алгоритм  $a_y$ , що  $a_y(x) = 1$ . Величина  $\delta_{a_1, \dots, a_i}$  характеризує інформаційну ефективність розпізнавання приналежності деякого об'єкту  $x$  до одного з класів  $H_0, H_1, \dots, H_{k-1}$  при умові, що  $x \in G_{a_1, \dots, a_i}$ .

Базове правило класифікації для структур дерев алгоритмів має наступний вигляд:

$$\rho_{a_1, \dots, a_i} = \psi_{a_1, \dots, a_i}^j$$

Чим більше величина  $\rho_{a_1, \dots, a_i}$ , тим вища ефективність правила класифікації.

Середня ефективність розпізнавання набору образів  $H_0, H_1, \dots, H_{k-1}$ , які задані даними НВ за допомогою алгоритмів розпізнавання  $\alpha_1, \alpha_2, \dots, \alpha_n$  оцінюється наступною величиною:

$$F_S(\alpha_1, \alpha_2, \dots, \alpha_n) = \sum_{a_1, \dots, a_i} \delta_{a_1, \dots, a_i} * \rho_{a_1, \dots, a_i}$$

Величина  $\tau_F$  характеризує загальну ефективність ФР  $F(a_1, \dots, a_i)$  для початкової НВ відносно деякого набору алгоритмів класифікації  $\alpha_1, \dots, \alpha_n$ , представляється в наступній формі:

$$\tau_F = \frac{m_F}{m}$$

Тут величина  $m_F$  – представляє кількість всіх входжень навчальних пар  $(x_j, f_R(x_j))$  в початкову НВ, які правильно класифікуються ФР  $F(a_1, \dots, a_i)$ .

Структура АДК (типу I) буде мати загальну конструкцію вигляду (Рис. 9), де кожний ярус такого дерева класифікації визначає етап побудови АДК у вигляді апроксимації поточним алгоритмом класифікації  $a_i$  певної частини НВ.

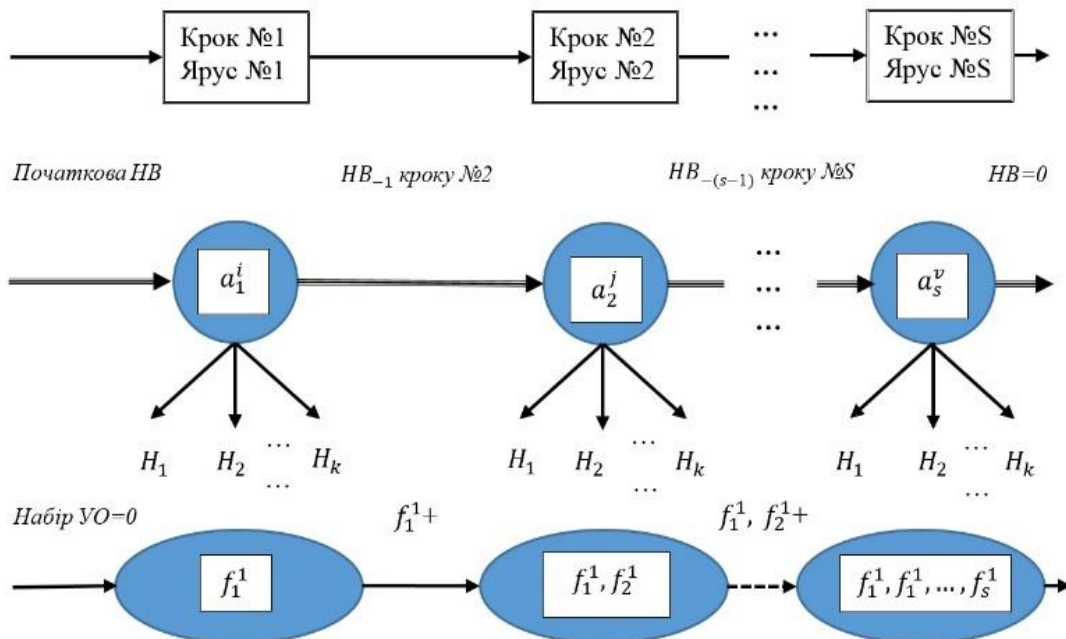


Рисунок 9 – Схема структури АДК типу I.

На кожному кроці генерації моделі АДК подається свій алгоритм  $\alpha_i$  класифікації та своя відповідна НВ (або підмножина початкової НВ). Початкова НВ у повному складі подається лише на першому кроці, далі з наступними етапами побудови дерева



класифікації потужність масиву даних НВ буде падати за рахунок набору побудованих узагальнених ознак (УО)  $f_j$ , які будуть відрізати (апроксимувати) певну частину даних початкової НВ. У залежності від структури схеми побудови АДК та особливостей поточного алгоритму  $\alpha_i$  на кожному кроці можливо генерувати більше однієї УО  $f_j$ .

У розділі вводяться критерії побудови моделі дерева алгоритмів – критерій зупинки процедури розгалуження  $K_{stop}$  та критерій відбору розгалуження  $W(a)$  для дерева класифікації, що будується:

1) Критерій зупинки  $K_{stop}$  процесу розгалуження типу (*boolean*) процедури побудови АДК, який полягає в перевірці потужності  $P_{pt}(NB)$  початкової вибірки наступного вигляду:

$$K_{stop} = \begin{cases} 0, & \text{if } P_{pt}(NB) = 0 \\ 1, & \text{if } P_{pt}(NB) > 0 \end{cases}.$$

2) Критерію розгалуження структури АДК у вигляді оцінки ефективності набору алгоритмів  $(\alpha_1, \alpha_2, \dots, \alpha_i)$ :

$$W(\alpha_i) = \frac{\frac{1}{P_{pt}(NB)} * \sum_{j=1}^k (T_{y3} + S_{y3} + \frac{E_{y3}}{S_{y3}})}{k}.$$

Уведені величини мають наступну інтерпретацію:

- $k$  – загальна кількість класів поточної задачі, які задані розбиттям  $R$  даних початкової НВ;
- $T_{y3}$  – характеризує загальний час (апаратний час), який витрачається на побудову поточної УО  $f_j$ ;
- $E_{y3}$  – інформаційна ємність (структурна складність) побудованої УО  $f_j$  на поточному кроці генерації моделі АДК;
- $S_{y3}$  – загальна кількість дискретних об'єктів  $x_i$  початкової НВ, які узагальнює (апроксимує) дана УО  $f_j$ ;
- $P_{pt}(NB)$  – потужність (об'єм) початкової НВ (або її фіксованої частини для поточного кроку схеми побудови АДК).

Схема побудови АДК першого типу полягає в наступному.

Етап 1. У бібліотеці алгоритмів інформаційної системи відбирається (в інтерактивному режимі, випадковим чином або після процедури відповідної оцінки якості, ефективності за даними початкової НВ) набір автономних алгоритмів класифікації та розпізнавання  $(\alpha_1, \alpha_2, \dots, \alpha_m)$ . Відбираються як самі алгоритми класифікації, так і їх кількість (величина  $m$ ) в залежності від умов та аспектів прикладної задачі.

Етап 2. Відібраний набір алгоритмів класифікації  $(\alpha_1, \alpha_2, \dots, \alpha_m)$  оцінюється та ранжується на основі функціоналу критерію розгалуження АДК за даними НВ у наборі відповідно до їх ефективності. Можливі два варіанти в залежності від алгоритмічної схеми побудови дерева класифікації:

а) Варіант, коли оцінка ефективності та ранжування набору алгоритмів класифікації  $(\alpha_1, \alpha_2, \dots, \alpha_m)$  проводиться лише один раз на даному етапі, а далі на кожному кроці побудови АДК фіксується для апроксимації даних наступний алгоритм  $\alpha_i$  початкової послідовності. Такий підхід значно економить апаратні

ресурси інформаційної системи, але негативно впливає на складність отриманої моделі дерева класифікації.

б) Варіант, коли оцінка ефективності та ранжування набору алгоритмів класифікації  $(\alpha_1, \alpha_2, \dots, \alpha_m)$  проводиться на кожному кроці побудови АДК за відповідними даними підмножин (частин) початкової НВ з метою оцінки та виявлення найбільш якісного (ефективного) алгоритму класифікації для даної частини НВ (кроку генерації АДК).

Такий підхід дозволяє за меншу кількість кроків завершити апроксимацію НВ та отримати більш економну конструкцію АДК у порівнянні з варіантом (а), однак потребує значно більше апаратних ресурсів інформаційної системи для другого етапу схеми побудови АДК та вимагає значної уваги та введення набору обмежень щодо початкового відбору послідовності алгоритмів класифікації  $(\alpha_1, \alpha_2, \dots, \alpha_m)$ .

Етап 3. Фіксується, як початкова вершина АДК – алгоритм класифікації  $\alpha_i$  найбільшої ефективності з відсортованого набору  $(\alpha_1, \alpha_2, \dots, \alpha_m)$ , та на вхід подається початкова НВ у вигляді послідовності навчальних пар. Алгоритм  $\alpha_i$  забезпечує генерацію однієї або декількох УО  $f_i$  першого ярусу (кількість УО, які генеруються на кожному кроці – задається параметрами схеми побудови АДК), які апроксимують певну частину НВ.

Етап 4. Обирається в якості вершини другого ярусу наступний за ефективністю алгоритм класифікації  $\alpha_i$  ранжованої послідовності  $(\alpha_1, \alpha_2, \dots, \alpha_m)$  та повторюється процедура побудови УО третього етапу з тією лише різницею, що на вхід подається вже обмежена НВ без навчальних пар, які апроксимуються УО вершиною першого ярусу і так далі. Процедура побудови АДК буде зводитися по повторення даного етапу для наступних за ефективністю алгоритму  $\alpha_i$  послідовності  $(\alpha_1, \alpha_2, \dots, \alpha_m)$ , постійного відсікання частин НВ та перевірки критерію зупинки розгалуження, якій фактично сигналізує про завершення процедури побудови моделі АДК та отримання на виході як дерева алгоритмів  $\alpha_i$  класифікації так і дерева узагальнених ознак  $f_i$ .

Зафіксовано, що алгоритмічне дерево першого типу не є єдиною можливою конструкцією (структурою) алгоритмів класифікації, які можна організувати у вигляді деревоподібної моделі розпізнавання. Структура АДК (тип II) (Рис. 10) складається з відповідних ярусів (на кожному з яких розташований відібраний алгоритм  $\alpha_i$  з побудованою УО), нумерація яких співпадає з послідовністю ранжованого за показником ефективності набору алгоритмів класифікації  $(\alpha_1, \alpha_2, \dots, \alpha_m)$ . Така схема структури АДК на виході не фіксує клас належності об'єкту класифікації, а остаточне рішення щодо віднесення до того чи іншого класу (на етапі тестування та роботи моделі АДК) приймається шляхом голосування після проходження об'єктом всіх вузлів, вершин даної логічної структури. В схемі АДК другого типу (Рис. 10) показане регулярне дерево класифікації, в якому на кожному ярусі розташовані однакові відібрані алгоритми класифікації  $\alpha_i$ . АДК генерує нерегулярне дерево УО, в якому на фіксованому ярусі логічної структури може розміщуватись різна кількість УО, які також можуть відрізнятися параметричною складністю та потужністю, тобто залежати від результатів попереднього кроку апроксимації даних НВ.

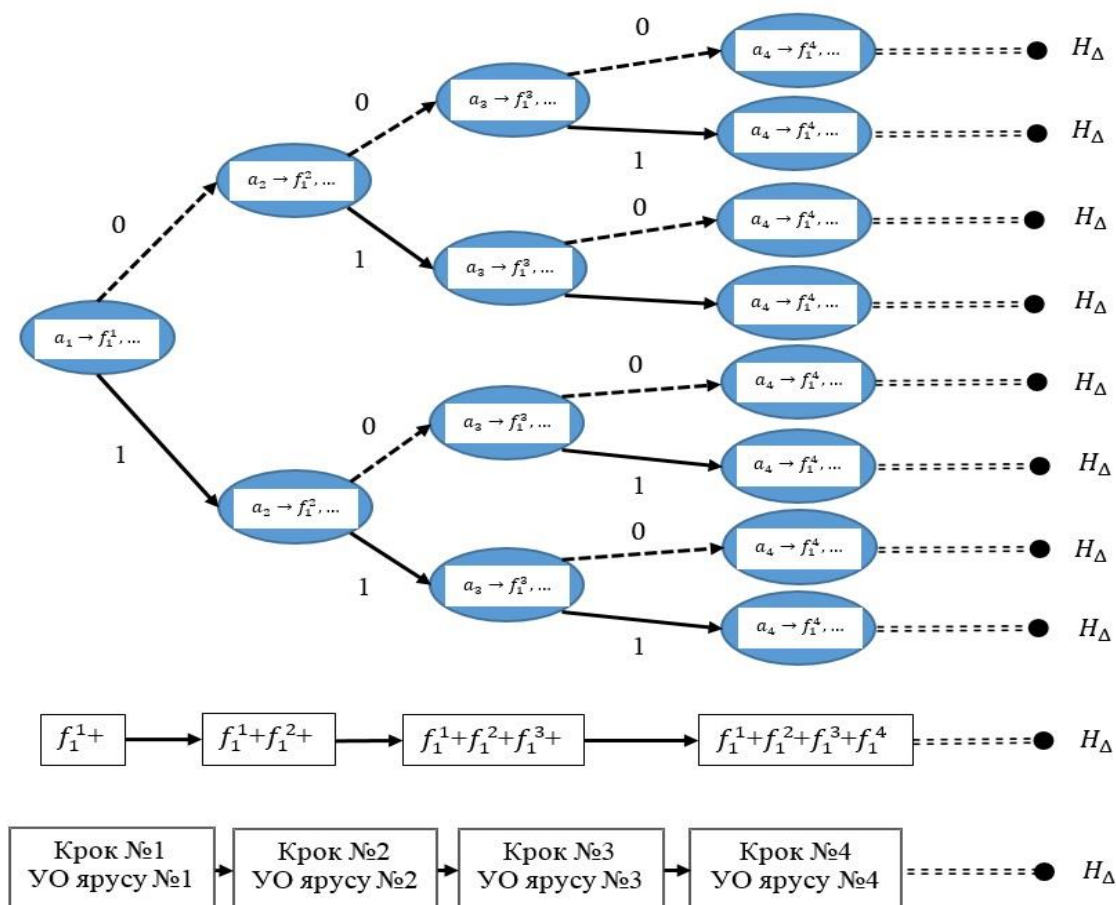


Рисунок 10 – Схема структури АДК типу II.

У розділі запропонований загальний інтегральний показник якості моделі АДК в наступній формі:

$$Q_{Main} = \frac{Fr_{All}}{V_{All} \cdot \sum_i p_i} \cdot e^{-\frac{Er_{All}}{M_{All}}}.$$

Тут набір параметрів  $p_i$  представляє собою найважливіші характеристики побудованого дерева класифікації, що оцінюється:

- $Er_{All}$  – загальна кількість помилок моделі АДК на масивах даних початкових тестової та навчальної вибірки;
- $M_{All}$  – загальна потужність масивів даних навчальної та тестової вибірки;
- $Fr_{All}$  – кількість вершин отриманої моделі АДК з результуючими значеннями  $f_R$ ;
- $V_{All}$  – загальна кількість усіх типів вершин в структурі моделі АДК;
- $O_{Uz}$  – загальна кількість узагальнених ознак, що використовуються в моделі дерева класифікації;
- $P_{All}$  – загальна кількість переходів між вершинами в структурі побудованої моделі дерева класифікації;
- $N_{Alg}$  – загальна кількість різних автономних алгоритмів класифікації  $a_i$ , що використовуються в моделі дерева класифікації.

У межах дослідження методи алгоритмічного дерева класифікації (методи АДК першого та другого типу) порівнювалися з методом повного ЛДК та обмеженого методу селекції елементарних ознак на даних задачі компонентного хімічного аналізу вмісту пального. Побудовані моделі АДК показали прийнятний результат (Табл. 2).

Таблиця 2 – Порівняльна таблиця моделей / методів дерев класифікації

№	Метод синтезу структури логічного дерева	Інтегральний показник якості моделі $Q_{Main}$	Загальна кількість помилок моделі на НВ та ТВ $Er_{All}$	Загальний час генерації дерева класифікації $T_{All}$
1	Метод повного ЛДК на основі селекції елементарних ознак	0,004789	2	264 с.
2	Модель ЛДК з одноразовою оцінкою важливості ознак	0,002263	3	232 с.
3	Обмежений метод побудови ЛДК	0,003181	4	212 с.
4	Метод алгоритмічного дерева (типу I)	0,005234	0	321 с.
5	Метод алгоритмічного дерева (типу II)	0,002941	0	359 с.
6	Метод повного ЛДК (алг. Ю.А. Василенка)	0,005032	0	282 с.
7	Обмежений метод побудови АДК	0,003043	5	238 с.
8	Схема С4.5 (повна структура ЛДК)	0,004812	0	249 с.
9	Випадковий ліс в схемі С5.0 (набори випадкових ЛДК неповної структури)	$(Q_{Main}^L)$ 0,005325	0	567 с.

Розроблено обмежений метод побудови АДК. Такий підхід синтезу моделі розпізнавання дозволяє досить ефективно регулювати складність, точність моделі дерева класифікації, що будується. Доцільним є його застосування в ситуаціях з обмеженнями щодо апаратних ресурсів інформаційної системи, обмеженнями точності та структурної складності моделі, обмеженнями на структуру, послідовність та глибину розпізнавання масиву даних НВ.

Схема обмеженого методу побудови моделей АДК полягає в наступному. АДК першого типу складається з ярусів, кожний з яких відповідає певному кроку (етапу) побудови (апроксимації даних початкової НВ) дерева класифікації. Для кожного алгоритму класифікації на тому чи іншому кроці апроксимації можна розрахувати його ефективність відносно робочих даних –  $(S/P_{pt}(NB^-))$ . Дана величина має бути більше або дорівнювати початковому обмеженню  $\delta$  (у деяких реалізаціях схеми АДК величина  $\delta$  може бути використана в якості критерію зупинки процедури розгалуження в структурі дерева класифікації, критерію попередньої обрізки). Тут  $S$  – загальна кількість помилок класифікації для фіксованого алгоритму на певному кроці (етапі генерації АДК), а  $P_{pt}(NB^-)$  – потужність підмножини початкової НВ, яка подається на вхід даного алгоритму на відповідному ярусі, рівні або етапі дерева класифікації, що будується. Тоді доцільно для кожного ярусу структури АДК, етапу побудови дерева класифікації розраховувати величини  $S_{a_1, \dots, a_i}$ , які характеризують

кількість всіх пар  $(x_i, f_R(x_i))$  з масиву початкової НВ, котрі не можуть бути апроксимовані послідовністю фіксованих алгоритмів класифікації  $a_1, \dots, a_i$ . Отже,  $S_{a_1, \dots, a_i}$  – це число всіх тих помилок класифікації, які здійснюється деякою послідовністю УО, побудованих на відповідних рівнях структури АДК для фіксованого набору алгоритмів розпізнавання та класифікації  $a_1, \dots, a_i$ . Тоді за даною схемою побудови обмеженої структури АДК вибирається лише той шлях або певні шляхи в конструкції дерева (в залежності від типів дерев класифікації можуть бути різні), для якого (яких) величина  $S_{a_1, \dots, a_i}$  буде по можливості максимальною. Добудовується шлях у структурі АДК з найбільшою кількістю помилок класифікації, причому наступна добудова та відбір вершин (алгоритмів класифікації – параметрів УО)  $a_1, \dots, a_i$  здійснюється тільки для цих шляхів (Рис.11).

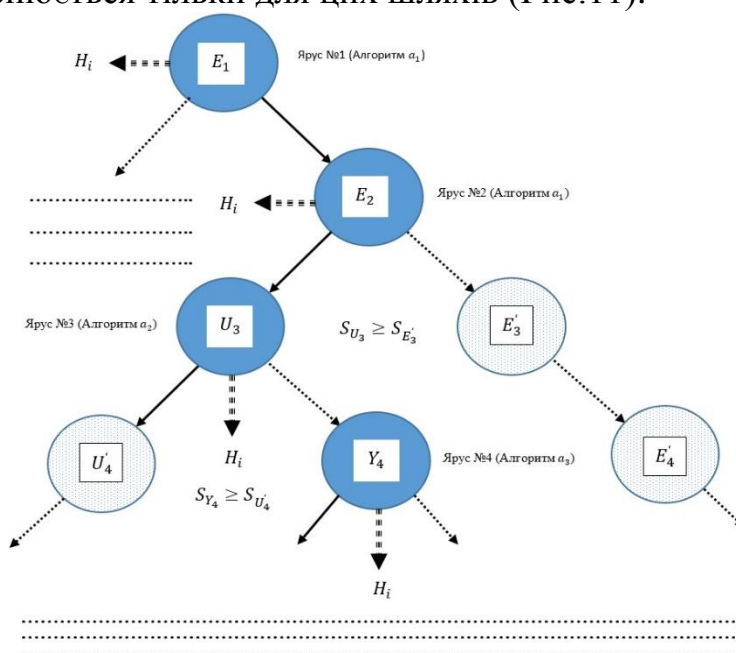


Рисунок 11 – Фрагмент структури обмеженого АДК.

У розділі досліджено питання збіжності процедури побудови моделей дерев класифікації методів ЛДК/АДК для умов слабого та сильного розділення класів початкової НВ, наведено відповідні числові оцінки. Загальна кількість усіх кінцевих вершин логічної структури (листя дерева розпізнавання) побудованої схеми класифікації буде однозначно визначати кінцеву потужність схеми методу дерева класифікації (моделей ЛДК/АДК).

Встановлено, що для умови слабого розділення класів у випадку ЛДК, якщо на кожному  $n$  – вому кроці відібрана елементарна ознака  $\varphi_n$  слабо розділяє множину (підмножину) об'єктів початкової НВ, то в цьому разі процес побудови дерева класифікації збігається відносно початкової НВ та закінчується не більше ніж за  $m - 1$  кроків, де  $m$  – кількість всіх навчальних пар початкової НВ. При умові сильного розділення класів множини об'єктів початкової НВ, яке має  $m$  повних ярусів, рівнів (тобто випадок, коли на  $i$  – товому ярусі стоять  $2^{i-1}$  вершин), має  $2^{m+1} - 1$  вершин – таким чином розпізнавання масиву початкової НВ при умові ( $y > 1$ ) за допомогою повного ЛДК відбувається не більш ніж за  $2^{m+1} - 1$  кроків, де  $m$  розраховується за допомогою виразу:

$$m = R\left(\frac{\log_2 n_1}{1 + \log_2 y}\right).$$

Для випадку слабого розділення класів початкової НВ у схемі АДК процес побудови дерева класифікації збігається відносно масиву даних НВ та закінчується не більше ніж за  $M$  кроків, де  $M$  – кількість всіх навчальних пар початкової НВ. У разі сильного розділення класів початкової НВ для схеми АДК, коли потужність побудованої УО (або набору УО) обмежена лише практичною можливістю самого алгоритму класифікації  $a_i$  та початковими параметрами НВ – схема (модель) АДК буде побудована за  $t$  кроків, де величина  $t$  визначається співвідношенням:

$$t \leq 2 * \frac{P_{pt}(НВ)}{z^{max} + z^{min}} = \frac{2M}{z^{max} + z^{min}}.$$

Для випадку обмеження щодо потужності синтезованих УО (не перевищення відповідної величини  $P$ ) – схема дерева класифікації (модель АДК) буде побудована за  $t$  кроків, де величина  $t$  визначається співвідношенням  $t \leq \frac{M}{P}$ .

У **сьомому розділі** розроблено програмну схему обчислення важливості ознак (груп ознак), яка дає змогу вирішувати широке коло практичних задач. Розглядається питання кодування початкової інформації НВ, один із шляхів вирішення якої полягає в схемі розбиття ознакового простору задачі на  $n$  – вимірні гіперпаралелепіеди так, щоби елементи (об'єкти) з різних класів не попадали в один і той самий паралелепіед (задача геометричного обмеження) з присвоєнням кожному з даних геометричних об'єктів фіксованого двійкового коду.

Розроблено комплексну програмну систему (ПС) – DeTree, дає можливість забезпечити ефективний механізм програмної побудови та фінального аналізу фіксованого ЛДК (набору ЛДК) за масивом початкових даних НВ. Програмний інструмент розділений на два проекти (компоненти) – *DeTreeBackend* та *DeTree*. Компонент *DeTreeBackend* представляється як бекенд, а *DeTree* як фронтенд, причому в даній схемі компонент *DeTreeBackend* забезпечує функціонал усіх базових обчислень (усіх схем обробки даних та менеджера пам'яті), додатково містить у собі набір класів та алгоритмів низькорівневого функціоналу. Компонент *DeTree* в свою чергу реалізує повний функціонал роботи з боку користувача (інтерактив) та забезпечує роботу довідкової та сервісної служб.

Послідовність кроків побудови структури дерева класифікації (моделі ЛДК) визначається наступним порядком дій:

1) Підготовчий етап початкового вибору, ініціалізація базових параметрів та налаштувань ПС під реалії конкретної прикладної задачі відповідно до умов генерації структури (моделі) ЛДК – критеріїв розгалуження та критеріїв зупинки побудови дерева класифікації.

2) Початковий етап валідації наборів вхідних даних поточної задачі (перевірка на коректність різних типів даних НВ та ТВ), вибір та перевірка на коректність основних режимів (параметрів) роботи процедури генерації дерева класифікації.

3) Етап забезпечення первинних процедур запиту, виділення та розподілу оперативної пам'яті системи під центральну вершину (*Root Node*), вузли та переходи структури дерева класифікації (робота менеджера пам'яті ПС).

4) Етап роботи рекурсивної процедури обрахунку набору величин інформативності (важливості) атрибутів (ознак)  $W(P_i)$ , відповідно до обраного автоматично або в інтерактивному режимі критерію розгалуження структури ЛДК.

5) Етап генерації структури ЛДК (вузлів дерева класифікації) відповідно до зафіксованих критеріїв розгалуження та зупинки побудови дерева класифікації.

6) Етап роботи процедури відсікання (оптимізації структури ЛДК) для мінімізації структурних компонентів (вузлів, блоків) побудованого дерева класифікації.

7) Етап фінальної перевірки основних параметрів побудованої структури ЛДК (моделі класифікації) в автоматичному або інтерактивному режимі в залежності від початкових налаштувань ПС.

Проведено оцінювання та перевірку ПС DeTree на даних НВ об'ємом  $\approx 600$  тис. об'єктів у двох апаратних конфігураціях. Встановлено, що побудовані схемою С5.0 набори правил (модель ЛДК) мають помітно нижчу частоту помилок порівняно з С4.5 та DeTree, причому мають однакову точність готової моделі, але набір правил в моделі С5.0 є незначно меншим. Відносно часу генерації структур ЛДК (та правил класифікації) зафіксовано, що алгоритм С4.5 набагато швидший за рахунок простоти схеми, достатньої оптимізації та обмежень на процедуру фінальної обрізки (*pruning*).

Так, витрати оперативної пам'яті схеми С5.0 зазвичай на порядок менше ніж у С4.5 при побудові набору правил (моделі ЛДК): було використано 230 МВ, а для С4.5 витрачено 4.3 GB, але все рівно більше ніж у ПС DeTree – 180 МВ.

Таблиця 3 – Порівняння схем побудови ЛДК за кількістю помилок класифікації, кількістю правил класифікації, та часом побудови ЛДК

Алгоритмічна схема	Загальна кількість помилок - $Er_{All}$	Загальна кількість правил класифікації - $Rt_{All}$	Загальний час генерації ЛДК - $T_{All}$
С4.5	7.2%	5420	Config. №1 – 34 с. Config. №2 – 42 с.
С5.0	6.3%	4845	Config. №1 – 186 с. Config. №2 – 230 с.
DeTree	6.7%	5028	Config. №1 – 102 с. Config. №2 – 129 с.

На основі проведеного аналізу встановлено, що на сьогоднішній день відомо десятки ПС для побудови різних типів моделей дерев класифікації (дерев рішень) у вигляді структур ЛДК та лише одна ПС, яка базується на концепції АДК, причому всі ці системи відрізняються прикладною спрямованістю задач що розв'язуються, методами та концептуальними засадами, різноманітним рівнем підтримки, причому багато з них знаходять у вільному (або частково вільному) доступі.

Таблиця 4 – Порівняння схем побудови ЛДК за фіксованою точністю, кількістю вузлів та часом побудови структури ЛДК

Алгоритмічна схема	Загальна кількість помилок - $Er_{All}$	Загальна кількість вузлів ЛДК - $Vt_{All}$	Загальний час генерації ЛДК - $T_{All}$
С4.5	6.8%	10167	Config. №1 – 57 с. Config. №2 – 43 с.
С5.0	6.8%	9201	Config. №1 – 62 с. Config. №2 – 51 с.
DeTree	6.7%	1012	Config. №1 – 50 с. Config. №2 – 47 с.

У роботі побудовано моделі дерев класифікації (структури АДК) на основі даних гідрографічного спостереження загального стану басейну річки Уж для виявлення ситуації червоної (паводкової) зони на основі поточних замірів постів спостережень.

Модель паводкового явища описується на основі 18 ознак (атрибутів), які мають різну природу та формуються на основі багаторічних спостережень басейну річки Уж. На основі набору представлених гідрографічних характеристик були побудовані моделі класифікації паводкових явищ для річки Уж за 19 річний період (1992 - 2010) у вигляді структур (моделей) АДК. Для побудови моделей дерев класифікації використовувалася ПС «Оріон III» для генерації автономних систем розпізнавання та класифікації, де алгоритмічна бібліотека системи нараховує 15 алгоритмів (методів та схем розпізнавання). Загальні параметри (характеристики) побудованих структур (моделей АДК) представлені в (Табл. 5).

У масиві навчаючої інформації переважали навчальні пари класу  $H_3$  (об'єкти ситуаційного стану нейтральної зони, зеленого маркеру) на другому місці зі значним відривом за кількістю знаходились навчальні пари класу  $H_2$  (об'єкти ситуаційного стану спостережної зони, жовтого маркеру) і на третьому місці – безпосередньо навчальні пари паводкових явищ (об'єкти червоного маркеру) – класу  $H_1$ . Потужність класу  $H_2$  незначно переважає потужність класу  $H_1$ , це пояснюється динамікою зміни паводкової ситуації в часі, яка може повертатися до нормального стану (нейтральної зони) – явищ класу  $H_3$ , а в більшості випадків переходить в кризовий стан (червону зону паводкового явища) – класу  $H_1$  (Рис. 12).

Таблиця 5 – Загальні параметри побудованих моделей АДК

Номер побудованої моделі АДК	Метод (алгоритм) побудови моделі АДК	Загальна кількість різних алгоритмів класифікації застосованих в АДК - $N_{All}$	Загальна кількість УО (наборів УО) в структурі АДК $O_{Uz}$	Загальна кількість всіх вершин (разом з результируючими) в структурі АДК	Загальний час генерації структури АДК поточної НВ
<i>Пост спостереження №1</i>					
1	Метод повного АДК (тип I)	5 обмеження на послідовне використання по одному алгоритму	42	84	686 с.
2	Метод повного АДК (тип II)	5 обмеження на кількість генерацій УО на один крок побудови АДК	43	98	712 с.
3	Обмежений метод АДК ( $Z=10$ )	1 алг. гіперсфер	35	71	406 с.
4	Обмежений метод АДК ( $Z=5$ )	1 алг. гіперкуба	58	117	839 с.
5	Обмежений метод АДК ( $Z=3$ )	1 алг. гіпереліпс	37	75	442 с.
6	Метод повного АДК (тип I)	2 алг. гіперсфер алг. гіперкуба	40	81	706 с.
<i>Пост спостереження №2</i>					
1	Метод повного АДК (тип I)	5 обмеження на послідовне використання по одному алгоритму	39	79	676 с.



2	Метод повного АДК (тип II)	5 обмеження на кількість генерацій УО на один крок побудови АДК	40	95	701 с.
3	Обмежений метод АДК (Z=10)	1 алг. гіперсфер	34	69	380 с.
4	Обмежений метод АДК (Z=5)	1 алг. гіперкуба	57	115	828 с.
5	Обмежений метод АДК (Z=3)	1 алг. гіперпаралелепіед	48	98	793 с.
6	Метод повного АДК (тип I)	2 алг. гіперсфер алг. гіперкуба	38	77	695 с.

Масив НВ складався з 8391 об'єктів (наборів відомої класифікації) для двох пунктів моніторингу на ділянці міста Ужгород, причому ефективність сконструйованої системи розпізнавання оцінювалася на тестовій вибірці об'єму 500 об'єктів на кожний з постів спостереження, масив ТВ представляв собою відокремлену частину початкової НВ.

Фрагмент основних результатів інтегральної оцінки якості побудованих структур (моделей) АДК, проведених експериментів – порівняльних тестів методів побудови моделей АДК (структур дерев класифікації) на масиві даних даної прикладної задачі, представлений в (Табл. 6).

У розділі зафіксовано, що побудовані дерева класифікації (моделі АДК) забезпечили необхідні якість та швидкість схем класифікації паводкових явищ річки Уж при достатньо компактній структурі самої конструкції дерева.

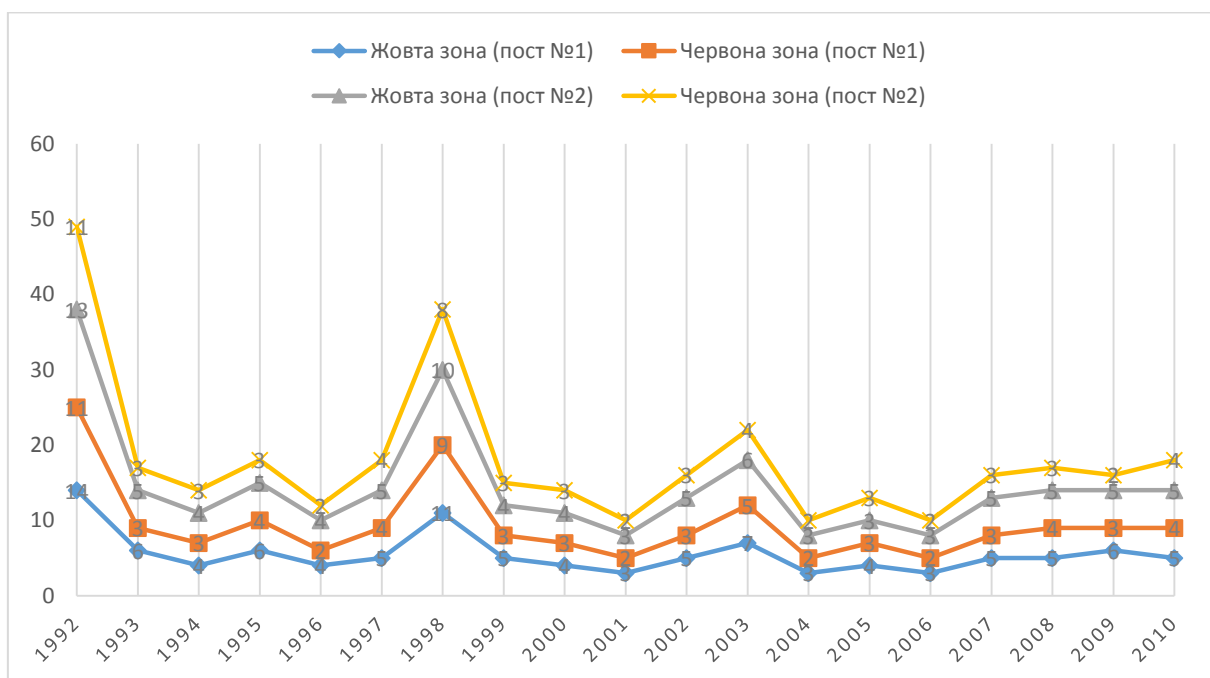


Рисунок 12 – Паводкові явища жовтої та червоної зони річки Уж в період 1992 – 2010 р.

Проведені в роботі практичні випробовування моделей АДК підтвердили працездатність математичного забезпечення та запропонованих методів та алгоритмів побудови дерев класифікації, розробленого програмного забезпечення.

Таблиця 6 – Порівняльна таблиця моделей класифікації (структур АДК) класифікації паводкових явищ річки Уж (постів №1-№2).

№	Метод (схема) синтезу структури (моделі) дерева класифікації (ЛДК/АДК)	Інтегральний показник якості моделі дерева класифікації $Q_{Main}$	Загальна кількість помилок моделі на НВ та ТВ $Er_{All}$
1	Метод повного АДК (тип I) (кількість алгоритмів – 5, обмеження на послідовне використання по одному алгоритму)	0,005821 0,005845	0 0
2	Метод повного АДК (тип II) (кількість алгоритмів – 5, обмеження на кількість генерацій УО на один крок побудови АДК)	0,004778 0,004712	0 0
3	Обмежений метод АДК (Z=10) (кількість алгоритмів – 1, алг. гіперсфер)	0,004464 0,004389	0 0
4	Обмежений метод АДК (Z=5) (кількість алгоритмів – 1, алг. гіперкуба)	0,004387 0,004228	12 13
5	Обмежений метод АДК (Z=3) (кількість алгоритмів – 1, алг. гіперпаралелепіпеда)	0,004256 0,004354	6 6
6	Обмежений метод АДК (Z=3) (кількість алгоритмів – 1, алг. гіпереліпса)	0,005582 0,005645	1 0
7	Метод повного АДК (тип I) (кількість алгоритмів – 2, алг. гіперсфер, алг. гіперкуба)	0,005790 0,005801	0 0

Набори незалежних алгоритмів класифікації, які були відібрані для генерації груп УО також підтвердили свою ефективність у межах даної прикладної задачі. Можливим шляхом подальших досліджень може бути розширення переліку алгоритмів класифікації в схемі АДК, а також додаткові умови та обмеження щодо генерації наборів УО для кожного кроку схеми дерева класифікації (структури АДК).

Це дає можливість рекомендувати використання даного підходу (концепції моделей АДК) та його програмної реалізації для широкого спектру прикладних задач розпізнавання в практичній площині.

**У додатках** наведено акти впровадження, копію свідоцтва про реєстрацію авторського права на твір (комп'ютерна програма), сертифікати підтвердження участі в міжнародних наукових програмах.

## ВИСНОВКИ

У дисертаційній роботі на основі проведених досліджень вирішено 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.

При цьому отримано такі результати:

1. Виконано аналіз сучасного стану досліджень, методів, алгоритмів побудови деревоподібних моделей класифікації (дерев рішень). Проведений аналіз показав велику сегментацію підходів відносно прикладних задач та відсутність єдиної

методології раціонального використання накопиченого потенціалу методів та алгоритмів класифікації.

2. Розроблено комплексний метод побудови деревоподібних моделей класифікації, який базується на поетапній апроксимації масиву початкових даних НВ набором відібраних і оцінених незалежних алгоритмів розпізнавання, що забезпечує побудову різнотипних моделей класифікації, їх універсальність та можливість роботи з великими масивами різнотипних даних.

3. Розроблено метод T – опорних множин, який полягає у відборі (фіксації) певного набору ознак разом зі своїми значеннями (правилом класифікації в структурі ЛДК) на основі інформації деякої початкової НВ, з можливістю наступної оцінки даних опорних множин за допомогою відповідних функціоналів, що забезпечує ефективний механізм представлення дискретних об'єктів для структур логічних дерев класифікації.

4. Розроблено метод знаходження подібності конструкцій логічних дерев у задачах мінімізації їх структур, який за рахунок подібних вершин в структурі дерева класифікації забезпечує ефективний механізм фінальної обрізки побудованої структури дерева класифікації.

5. Розроблено методи оптимального розташування змінних (атрибутів, міток) конструкції ЛДК, які в більшості випадків дають оптимальне логічне дерево відносно його структури, що забезпечує можливість якісно провести процедуру фінальної обрізки побудованої структури ЛДК.

6. Розроблено метод обчислення впливу процедури перестановки ярусів у структурі регулярного логічного дерева на його загальну складність, що забезпечує зменшення складності його структури майже в 3 рази.

7. Розроблено методи побудови структур АДК різних типів (обмежених структур АДК), де отримані дерева класифікації складаються з різних методів, алгоритмів розпізнавання, в свою чергу представляють собою нові схеми класифікації. Це дає змогу розширити область застосування, будувати моделі класифікації, точність яких можна регулювати в процесі побудови, або будувати системи з наперед заданою точністю, раціонально використовувати вже накопичений потенціал методів та алгоритмів теорії розпізнавання. Для обмеженої схеми синтезу алгоритмічних дерев класифікації швидкість синтезу моделі в середньому зростає більше чим на 20%.

8. Встановлено питання збіжності процедури побудови моделей дерев класифікації (запропонованих в дослідженні структур АДК) для умов слабого та сильного розділення класів початкової НВ. Це забезпечує отримання параметрів моделі АДК максимальної складності в умовах поточної задачі.

9. Розроблено програмний інструментарій побудови структур ЛДК/АДК (моделей дерев класифікації різних типів), для розв'язання широкого спектру різнотипних прикладних задач розпізнавання дискретних об'єктів, автоматизації процедури синтезу моделей дерев класифікації різних типів (структур логічних та алгоритмічних дерев класифікації).

### СПИСОК ОСНОВНИХ ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Повхан І. Ф. Логічні дерева класифікації в задачах штучного інтелекту: монографія. Ужгород: Поліграфцентр - Ліра, 2019. 276 с.
2. Повхан І. Ф. Питання структурної складності для випадку регулярного логічного дерева. *Scientific and technical progress in European countries and the contribution of higher education institutions* : collective monograph. Riga : Izdevnieciba «Baltija Publisher», 2020. 308 p.
3. Povhan I. F. Logical recognition tree construction on the basis a step-to-step elementary attribute selection. *Radio Electronics, Computer Science, Control*. 2020. № 2. P. 95–106.
4. Povkhan I. F. The general concept of the methods of algorithmic classification trees. *Radio Electronics, Computer Science, Control*. 2020. № 3. P. 108–121.
5. Povhan I. F. Limited method for the case of algorithmic classification tree. *Radio Electronics, Computer Science, Control*. 2020. № 4. P. 106-118.
6. Povkhan I. Classification models of flood-related events based on algorithmic trees. *Eastern-European Journal of Enterprise Technologies*. 2020. Vol. 6, No 4 (108). P. 58-68. DOI: <https://doi.org/10.15587/1729-4061.2020.219525>.
7. Povhan I. Logical classification trees in recognition problems. *Kwartalnik Naukowo-Techniczny: Informatyka Automatyka Pomiary w gospodarce o ochronie srodowiska*. Krakow, 2020. № 2. P. 12–16. DOI: <http://doi.org/10.35784/iapgos.927>.
8. Povhan I. General scheme for constructing the most complex logical tree of classification in pattern recognition discrete objects. *Електроніка та інформаційні технології* : зб. наук. пр. Lviv, 2019. Vol. 11. С. 73–80.
9. Povhan I. Generation of elementary signs in the general scheme of the recognition system based on the logical tree. *Електроніка та інформаційні технології* : зб. наук. пр. Lviv, 2019. Vol. 12. С. 20–29.
10. Povhan I. Question of the optimality criterion of a regular logical tree based on the concept of similarity. *Електроніка та інформаційні технології* : зб. наук. пр. Lviv, 2020. Vol. 13. С. 19–27. DOI: <https://doi.org/10.30970/eli.13.2>
11. Повхан І. Ф. Задача загальної оцінки складності максимального побудованого логічного дерева класифікації. *Вісник Національного технічного університету «Харківський політехнічний інститут»*. Серія: Інформатика та моделювання. 2019. №13 (1338). С. 104–117.
12. Повхан І. Ф. Питання побудови деревоподібних моделей розпізнавання образів. *Вісник Національного технічного університету «Харківський політехнічний інститут»*. Серія: Інформатика та моделювання. 2019. № 28 (1353). С. 39–57.
13. Повхан І. Ф. Питання гнучкості логічних дерев класифікації в задачах розпізнавання образів. *Технічні науки та технології*. 2019. № 3 (17). С. 131–140.
14. Повхан І. Ф. Метод побудови алгоритмічного дерева другого типу на основі апроксимації навчальної вибірки набором алгоритмів класифікації. *Технічні науки та технології* : наук. журн. Чернігів, 2020. № 2 (20). С. 126–139.
15. Повхан І. Ф. Питання синтезу дискретних зображень в задачах розпізнавання образів. *Вісник Вінницького політехнічного інституту* : наук. журн. Вінниця, 2020. № 4 (151). С. 50 – 57.

16. Повхан І. Ф. Моделі алгоритмів розпізнавання у вигляді логічних дерев класифікації. *Збірник наукових праць Національного університету кораблебудування імені адмірала Макарова*. 2019. № 1 (475). С. 156–163.
17. Повхан І. Ф. Питання оцінки ефекту перестановки ярусів логічного дерева максимальної складності для бінарного випадку. *Збірник наукових праць Національного університету кораблебудування імені адмірала Макарова*. 2020. № 2 (480). С. 99–107.
18. Повхан І. Ф. Проблема функціональної оцінки навчальної вибірки в задачах розпізнавання дискретних об'єктів. *Вчені записки Таврійського національного університету імені Вернадського. Серія: Технічні науки*. 2018. Т. 29 (68), № 6. С. 217–222.
19. Повхан І. Ф. Поняття функції та алгебраїчної схеми розпізнавання в задачах класифікації дискретних об'єктів. *Вчені записки Таврійського національного університету імені Вернадського. Серія: Технічні науки*. 2019. Т. 30 (69), № 2. С. 171–177.
20. Повхан І. Ф. Задача апроксимації вибірки дискретних наборів геометричними об'єктами. *Вчені записки Таврійського національного університету імені Вернадського*. Київ, 2019. Т. 30 (69), № 3. С. 136–142.
21. Повхан І. Ф., Лавер В. О. Алгоритми побудови логічних дерев класифікації в задачах розпізнавання образів. *Вчені записки Таврійського національного університету імені Вернадського*. Київ, 2019. Т. 30 (69), № 4. С. 100–106.
22. Повхан І. Ф. Особливості випадкових логічних дерев класифікації в задачах розпізнавання образів. *Вчені записки Таврійського національного університету імені Вернадського. Серія: Технічні науки*. 2019. Т. 30 (69), № 5. С. 138–143.
23. Повхан І. Ф. Питання представлення дискретних зображень в задачах розпізнавання образів. *Вчені записки Таврійського національного університету імені Вернадського*. Київ, 2019. Т. 30 (69), № 6. С. 154–159.
24. Повхан І. Ф. Питання однозначного покриття зображень прямокутниками в задачах розпізнавання образів. *Вчені записки Таврійського національного університету імені Вернадського*. Київ, 2020. Т. 31 (70), № 1. С. 119–124.
25. Василенко Ю. А., Ващук Ф. Г., Повхан І. Ф., Повхан Л. С. Вычисление важности дискретных признаков (анализ некоторых подходов). *Восточно-европейский журнал передовых технологий*. 2010. №5/4 (47). С. 71–75.
26. Василенко Ю. А., Ващук Ф. Г., Повхан І. Ф. Проблема оцінки складності логічних дерев розпізнавання та загальний метод їх оптимізації. *Восточно-европейский журнал передовых технологий*. 2011. № 6/4 (54). С. 24–28.
27. Василенко Ю. А., Ващук Ф. Г., Повхан І. Ф. Загальна оцінка мінімізації деревоподібних логічних структур. *Восточно-европейский журнал передовых технологий*. 2012. № 1/4 (55). С. 29–32.
28. Повхан І. Ф. Особливості синтезу узагальнених ознак при побудові систем розпізнавання за методом логічного дерева. *Інформаційні технології та комп'ютерне моделювання ІТКМ-2019* : матеріали міжнар. наук.-практ. конф., Івано-Франківськ – Яремче, 20–25 трав. 2019. Івано-Франківськ, 2019. С. 169–174.

29. Povhan I. Generation of classification schemes in the form of logical trees in discrete object recognition problems. *Modern Problems of Mathematical Modeling, Automated Control and Information Technologies MCIT-2019* : International Scientific and Practical Conference, Rivne, November 14–16, 2019. Rivne, 2019. P. 195–200.
30. Povhan I. Designing of recognition system of discrete objects. *Proceedings of the «2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP)»*, August 23–27, 2016, Lviv, Ukraine. Lviv, 2016. P. 226–231.
31. Povkhan I., Lupei M. The algorithmic classification trees. *Proceedings of the «2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP)»*, August 21–25, 2020, Lviv, Ukraine. Lviv, 2020. P. 37–44.
32. Lupei M., Mitsa A., Povkhan I., Sharkan V. Determining the eligibility of candidates for a vacancy using artificial neural networks. *Proceedings of the «2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP)»*, August 21–25, 2020, Lviv, Ukraine. Lviv, 2020. P. 18–23.
33. Povkhan I. A constrained method of constructing the logic classification trees on the basis of elementary attribute selection. *CEUR Workshop Proceedings: Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020)*, Zaporizhzhia, Ukraine, April 15–19, 2020. Zaporizhzhia, 2020. Vol. 2608. P. 843–857. URL: CEUR-WS.org.
34. Василенко Ю. А., Повхан І. Ф. Автоматизація побудови схем класифікації. *Теорія прийняття рішень* : праці IV Міжнародної школи – семінару, Ужгород, 29 вересня – 4 жовтня 2008 р. Ужгород, 2008. С. 137-138.
35. Василенко Ю. А., Василенко Е. Ю., Бунда В. В., Бунда С. О., Лавер О. Г., Повхан І. Ф. Универсальный подход к анализу и обработке специальной информации. *Інновації в навчальному процесі вищих навчальних закладів: міжнародний та національний досвід* : зб. наук. ст. за матеріалами XV міжнар. наук.-практ. конф., Сніна (Словацька Республіка), 6–9 листоп. 2007 р. Ужгород : Ліра, 2008. С. 99–106.
36. Василенко Ю. А., Василенко Е. Ю., Бунда В. В., Повхан І. Ф., Чедреки Я. Н. Математическое конструирование распознающих систем для дискретных объектов. *Лісабонська стратегія як визначальний чинник європейської інтеграції в галузі освіти і науки* : матеріали XVI Міжнар. наук.-практ. конф., Ужгород – Гирляни, 6–9 трав. 2008 р. Ужгород : Ліра, 2008. С. 79–83.
37. Василенко Ю. А., Завілопуло А. М., Повхан І. Ф., Повхан Л. С. Синтез систем розпізнавання на основі схем-агентів. *Інститут електронної фізики Національної академії наук України – 2011*: матеріали міжнар. конф. молодих учених, Ужгород, 24–27 трав. 2011 р. Ужгород : Мистецька Лінія, 2011. С. 187–188.
38. Василенко Ю. А., Ващук Ф. Г., Повхан І. Ф. The importance of discrete signs. *Перспективні шляхи й напрями вдосконалення освітньої системи у світлі Болонського процесу* : зб. наук. доп. за матеріалами XX Міжнар. наук.-практ. конф., Ужгород (Україна) – Кошице (Словацьчина) – Мішкольц (Угорщина), 16–19 листоп. 2010 р. Ужгород : ЗакДУ, 2011. Вип. 2(21), Ч. 1. С. 46–56.
39. Василенко Ю. А., Ващук Ф. Г., Повхан І. Ф. Автоматизация построения систем классификации на основе схем-агентов. *Математическое моделирование, оптимизация и информационные технологии* : материалы 3-й междунар. конф., Кишинев, 19–23 марта 2012 г. Кишинев : Академия транспорта, информатики и коммуникаций, 2012. С. 444–446.

40. Повхан І. Ф. Побудова систем прогнозування економічних явищ на основі концепції логічного дерева класифікації. *Математичні методи, моделі та інформаційні технології в економіці* : матеріали VI міжнародної науково-методичної конференції, Чернівці, 18-19 квітня 2019 р. Чернівці : Друк –Арт, 2019. С. 134-136.

41. Повхан І. Ф. Загальна схема алгоритму усунення помилок розпізнавання в логічних деревах класифікації. *Technical Sciences: History, The Present Time, The Future, Eu Experience (informatics and cybernetics, electronics, radio engineering and communications, automation and computer technology, mechanic engineering, transport)*: Proceedings of the international scientific and practical conference, Wloclawek (Republic of Poland), September 27–28, 2019. Wloclawek, 2019. P. 42–45.

42. Повхан І. Ф. Проблематика методів логічних дерев класифікації в задачах розпізнавання. *Science, engineering and technology: global and current trends*: Proceedings of the international scientific and practical conference, Prague (Czech Republic), December 27–28, 2019. Prague, 2019. P. 28–32.

43. Повхан І. Ф. Модульна концепція побудови дерев класифікації. *Цифрова економіка та інформаційні технології* : матеріали міжнар. наук.-практ. конф., Київ, 15–16 квіт. 2020 р. Київ, 2020. С. 87–90.

44. Повхан І. Ф. Методи логічних дерев класифікації в задачах штучного інтелекту. *Priority directions of science development*: Abstracts of II International Scientific and Practical Conference, Lviv (Ukraine), 25–26 November, 2019. Lviv, 2019. P. 213–218.

45. Повхан І. Ф. Питання представлення неповністю визначених багатозначних логічних функцій у вигляді логічних дерев в задачах класифікації. *Topical issues of the development of modern science*: Abstracts of IV International Scientific and Practical Conference, Sofia (Bulgaria), 11–13 December, 2019. Sofia (Bulgaria), 2019. P. 390–396.

46. Повхан І. Ф. Модифікований метод побудови дерев класифікації. *Комплексне забезпечення якості технологічних процесів та систем* : матеріали X міжнар. наук.-практ. конф., Чернігів, 29–30 квіт. 2020 р. Чернігів, 2020. Т. 2. С. 167–169.

47. Повхан І. Ф. Питання загальної складності процедури побудови логічного дерева. *Проблеми інформатики та моделювання* : матеріали XX наук.-техн. конф., Харків-Одеса, 16–21 верес. 2020 р. Харків : НТУ «ХПІ», 2020. С. 68–74.

48. Povkhan I., Lupei M., Kliar M., Laver V. The issue of efficient generation of generalized features in algorithmic classification tree methods. *Data Stream Mining and Processing (DSMP 2020)*: Proceedings of the Third International Conference, Lviv, Ukraine, August 21–25, 2020. Cham: Springer, 2020. P. 98–113.

## АНОТАЦІЇ

**Повхан І.Ф. Методи та принципи побудови дерев класифікації дискретних об'єктів для інтелектуального аналізу даних.** – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора технічних наук за спеціальністю 05.13.23 – «Системи та засоби штучного інтелекту». – Національний університет «Львівська політехніка», м. Львів, 2021.

Дисертаційна робота присвячена дослідженню особливостей логічних та алгоритмічних дерев класифікації, деяких питань їх застосування в задачах інтелектуального аналізу даних, розпізнавання образів шляхом створення ефективних моделей класифікації та аналізу великих масивів даних, універсальних систем розпізнавання дискретних об'єктів.

Проаналізовано та досліджено деревоподібні моделі класифікації різних типів. Розроблено комплексний метод побудови деревоподібних моделей класифікації, який базується на поетапній апроксимації масиву початкових даних НВ набором відібраних та оцінених незалежних алгоритмів розпізнавання. Розроблено метод Т – опорних множин, який полягає у відборі, фіксації набору ознак разом зі своїми значеннями – класифікаторів в структурі ЛДК. Розроблено моделі та методи побудови структур АДК двох типів, де отримані дерева класифікації складаються з різних алгоритмів та методів розпізнавання й в свою чергу представляють собою нові алгоритми, схеми класифікації. Розроблено модель та метод побудови обмежених структур АДК, які спрямовані на добудову лише тих шляхів, ярусів конструкції дерева класифікації, де є найбільша кількість помилок усіх типів класифікації. Розроблено методи знаходження подібності конструкцій логічних дерев в задачах мінімізації їх структур. Здійснено числову оцінку впливу процедури обрізки побудованого ЛДК – перестановки ярусів, рівнів, блоків структури ЛДК в конструкції регулярного логічного дерева на його загальну складність. Здійснено оцінку збіжності процедури побудови моделей дерев класифікації, запропонованих в дослідженні структур АДК для умов слабого та сильного розділення класів початкової НВ.

*Ключові слова:* аналіз даних, системи розпізнавання, класифікації об'єктів, дерева рішень, набір атрибутів, критерій розгалуження, алгоритмічне дерево класифікації.

**I.F. Povkhan. Methods and principles of constructing discrete object classification trees for data mining.** –Qualifying scientific work on the rights of manuscripts.

Doctor of technical science thesis, specialty: 05.13.23 – "Artificial intelligence systems and means". – Lviv Polytechnic National University, Lviv, 2021.

Present thesis is devoted to studying the specific features of both logical and algorithmic classification trees, some issues of their use in the problems of intelligence data analysis and image recognition by creating the efficient models of classification and analysis of large and super-large data arrays and universal systems of discrete object recognition.

Tree-like classification models of various types are analyzed and studied. The concept of constructing tree-like classification models based on a step-by-step approximation of an array of initial TS data with a set of selected and evaluated independent recognition algorithms is proposed. The concept of T – reference sets is proposed, which consists in selecting and fixing a set of features together with their classifier values in the LCT structure. Models and methods for constructing ACT structures of two types are developed, where the resulting classification trees consist of various algorithms and recognition methods, which in turn represent new algorithms and classification schemes. A model and method for constructing bounded ACT structures are developed, which are aimed at completing only those paths, tiers of the classification tree structure where there are the largest number of errors of all classification types. Methods for finding the similarity of logical tree structures



in problems of minimizing their structures are developed. A numerical estimate of the impact of the pruning procedure of the constructed LCT – rearrangement of tiers, levels, and blocks of the LCT structure in the construction of a regular logical tree on its overall complexity is given. The convergence of the procedure for constructing classification tree models proposed in the study of ACT structures for conditions of weak and strong separation of initial TS classes is estimated.

*Keywords:* data analysis, object classification/recognition systems, decision trees, attribute set, branching criterion, algorithmic classification tree.

**Повхан И. Ф. Повхан И. Ф. Методы и принципы построения деревьев классификации дискретных объектов для интеллектуального анализа данных.** – Квалификационная научная работа на правах рукописи.

Диссертация на соискание ученой степени доктора технических наук по специальности 05.13.23 «Системы и средства искусственного интеллекта». – Национальный университет «Львовская политехника», Львов, 2021.

Диссертационная работа посвящена исследованию особенностей логических и алгоритмических деревьев классификации, некоторых вопросов их применения в задачах интеллектуального анализа данных, распознавания образов путем создания эффективных моделей классификации и анализа больших массивов данных, универсальных систем распознавания дискретных объектов.

В работе исследована и решена актуальная научно-прикладная проблема развития теории анализа и синтеза деревьев решений, разработки моделей, методов, прикладного инструментария интеллектуального анализа данных на основе логических и алгоритмических деревьев классификации с большей точностью, уменьшенной сложностью моделей и повышенной эффективностью классификации дискретных объектов.

Проанализированы и исследованы древовидные модели классификации различных типов. Предложен комплексный метод построения древовидных моделей классификации, который базируется на поэтапной аппроксимации массива начальных данных набором отобранных и оцененных независимых алгоритмов распознавания.

Предложен метод Т – опорных множеств, который заключается в отборе, фиксации набора признаков вместе со своими значениями – классификаторов в структуре ЛДК. Разработаны модели и методы построения структур АДК двух типов, где полученные деревья классификации состоят из различных алгоритмов и методов распознавания и в свою очередь представляют собой новые алгоритмы, схемы классификации.

Разработана модель и метод построения ограниченных структур АДК, которые направлены на достройку только тех путей, ярусов конструкции дерева классификации, где есть наибольшее количество ошибок всех типов классификации. Разработаны методы нахождения сходства конструкций логических деревьев в задачах минимизации их структур. Дана числовая оценка влияния процедуры обрезки построенного ЛДК – перестановки ярусов, уровней, блоков структуры ЛДК в конструкции регулярного логического дерева на его общую сложность. Дана оценка сходимости процедуры построения моделей деревьев классификации, предложенных

в исследовании структур АДК для условий слабого и сильного разделения классов начальной ОВ.

Разработан программный инструментарий построения структур ЛДК/АДК (моделей деревьев классификации различных типов), который позволяет решать широкий спектр разнотипных прикладных задач распознавания образов.

По результатам диссертационного исследования проведено экспериментальное моделирование, апробация разработанных моделей и методов на практических задачах, а определения эффективности предложенных в работе методов построения АДК осуществлено на основе интегральных показателей качества моделей.

*Ключевые слова:* анализ данных, системы распознавания, классификация объектов, деревья решений, набор атрибутов, критерий ветвления, алгоритмическое дерево классификации.

Підписано до друку 21.01.2021 р. Формат 60x90/16.  
Папір офсетний. Ум. друк. арк. 1,9.  
Наклад 100. Зам. 0179.

Видано та віддруковано в ТОВ "Поліграфцентр "Ліра":  
88000, м. Ужгород, вул. Митрака, 25

Свідоцтво про внесення до Державного реєстру видавців,  
виготівників і розповсюджувачів видавничої продукції  
Серія ЗТ №24 від 7 листопада 2005 року.