

Повхан І.Ф.

ДВНЗ «Ужгородський національний університет»

Сароз В.Я.

ДВНЗ «Ужгородський національний університет»

Легеца А.В.

ДВНЗ «Ужгородський національний університет»

МОДИФІКОВАНИЙ МЕТОД ДЕРЕВА АЛГОРИТМІВ ДРУГОГО ТИПУ В ЗАДАЧАХ КЛАСИФІКАЦІЇ

В статті побудований ефективний метод синтезу дерев класифікації за фіксованою початковою інформацією для задачі розпізнавання геологічних даних. Побудоване алгоритмічне дерево класифікації другого типу буде безпомилково класифікувати всю навчальну вибірку, за якою побудована схема класифікації, мати високу інтерпретабельність, просту структуру та складатися з автономних алгоритмів класифікації, розпізнавання в якості вершин граф-схеми. Розроблений метод побудови дерев алгоритмів дозволяє працювати з навчальними вибірками великого об'єму різномісної інформації дискретного типу, забезпечує високу точність схеми, раціонально використовує апаратні ресурси системи в процесі генерації кінцевої схеми класифікації, дозволяє будувати моделі з наперед заданою точністю. В роботі пропонується модульна схема побудови класифікаторів у вигляді структур дерев класифікації, дерев алгоритмів, яка дозволяє обійти обмеження традиційних методів дерев рішень. Модифікований метод дерева алгоритмів другого типу дозволяє значно зменшити обчислювальну складність моделі класифікації в порівнянні з деревами алгоритмів інших структур та забезпечити необхідну, регульовану точність фінального класифікатора. Пропонується підхід синтезу нових алгоритмів розпізнавання на основі бібліотеки вже відомих алгоритмів та методів теорії розпізнавання. Такий підхід до синтезу моделей класифікації дозволить комплексно будувати нові класифікатори, моделі класифікації на основі модульного принципу. На базі запропонованого метода дерева алгоритмів другого типу побудований набір моделей, які забезпечили ефективну класифікацію масиву геологічних даних. Концепція дерев алгоритмів різних типів дозволяє ефективно працювати в задачах класифікації з даними довільних шкал, де інформація задається в звичній для нас природній формі. Побудовані моделі дерев класифікації забезпечили відсутність помилок на даних навчальної та тестової вибірки, підтвердили працездатність підходу дерев алгоритмів другого типу.

Ключові слова: дерево класифікації, алгоритмічне дерево, дискретний об'єкт, алгоритм розпізнавання, ознака.

Постановка проблеми. Класифікація і розпізнавання дискретних об'єктів представляють собою центральні проблеми в сфері штучного інтелекту. Дані задачі відрізняються великою різноманітністю, різним ступенем структурної складності і значною застосовністю в багатьох секторах економічної і соціальної діяльності людини. У таких дисциплінах, як геологія, де проблеми класифікації вирішуються за допомогою складних інформаційних систем, важливість та інтенсивність досліджень у цій галузі добре задокументовані [1–10]. Ці задачі класифікації вимагають розробки та декомпозиції математичних моделей, адаптованих до конкретних досліджуваних систем. В даний час в області штучного

інтелекту відсутній універсальний прикладний підхід, здатний вирішити весь спектр цих складних проблем. Однак з'явилося кілька широко застосовуваних теорій та методологій, серед яких нейронні мережі займають особливе місце завдяки своїй універсальності у вирішенні широкого спектру задач класифікації [11–14]. У практичних сценаріях спеціально налаштовані штучні нейронні мережі часто перевершують традиційні алгоритми та усталені моделі дерев рішень, такі як методи градієнтного бусту, особливо в задачах, пов'язаних з неструктурованими даними, дискретними наборами зображень або текстовим вмістом. І навпаки, при роботі зі структурованими наборами даних, що містять великі обсяги

масивних дискретних даних, які демонструють різні функціональні простори, методи та алгоритми, засновані на дереві рішень, демонструють чіткі переваги [15]. Як правило, деревоподібні моделі класифікації полегшують ефективну обробку даних різного масштабу, представляючи вхідну інформацію в притаманному їй вигляді. Численні сучасні стратегії і концепції спрямовані на розробку систем розпізнавання (СР) і класифікацій з використанням логічних/алгоритмічних моделей деревовидних класифікацій (структур ЛДК/АДК). Зростаючий інтерес до деревоподібних граф-схемних представлень класифікаторів обумовлений їх численними корисними властивостями [16]. Однією з перспективних областей застосування деревоподібної моделі класифікації, особливо в області алгоритмічних дерев, є класифікація геологічної інформації [22].

Аналіз останніх досліджень і публікацій.

Дана робота зосереджується на теорії дерев рішень фіксованого типу з акцентом на дерева алгоритмів та класифікацію дискретних об'єктів [14, 23, 25]. Зафіксуємо, що дослідження [20] наголошує на тому, що правила класифікації та схеми прийняття рішень, отримані з будь-якого методу або алгоритму розгалуженого вибору ознак, мають деревоподібну логічну структуру. Типовий класифікатор дерева прийняття рішень містить організовану послідовність вузлів, ознак та атрибутів, структурованих за шарами або рівнями, кожен з яких встановлюється на певному етапі синтезу дерева класифікації [15].

Суттєвою проблемою, виявленою в роботі [18], є ефективна побудова деревоподібних структур розпізнавання, які можуть приймати форму графів алгоритмів (АДК-структур). Отже, методологія дерева рішень полегшує створення ефективних класифікаторів, заснованих на модульному принципі, використовуючи добре відомі алгоритми розпізнавання [19–21]. Дослідження [14] аналізує фундаментальні питання, пов'язані з побудовою деревоподібних структур дерев рішень у випадку коли ознаки є малоінформативними, включаючи їх набори та комбінації. У галузі аналізу даних інваріантна здатність структур ЛДК/АДК виконувати одновимірне розгалуження дозволяє аналізувати вплив, важливість та якість окремих змінних. Ця здатність необхідна для управління різними типами змінних як наборами предикатів. Постійною проблемою, пов'язаною з методами та структурами дерева рішень, є оцінка якості та ефективності вершин (узагальнених ознак), які служать автономними алгоритмами класифікації

[15]. Методи логічних дерев класифікації широко використовуються в аналізі даних з метою синтезу ефективних моделей, які передбачають значення цільової змінної на основі вихідного набору даних, відформатований як структурована навчальна вибірка [19]. З прикладної точки зору, для задач класифікації використовуються численні методи та алгоритми, засновані на концепції дерева рішень – однак особливою популярністю користуються C4.5/C5.0, CART. Методи C4.5/C5.0 використовують теоретико-інформаційний критерій для вибору вузла або вершини, тоді як алгоритм CART спирається на індекс Джині, який оцінює відносні відстані між розподілами класів у межах метрики навчального набору [20, 21]. Набір методів та алгоритмів для вибору розгалужених ознак (структур АДК) базується на оптимальному наближенні початкового навчального набору з використанням ранжованого ряду алгоритмів класифікації [22]. Ключовою проблемою в методах ЛДК/АДК, як відзначено в [23], є вибір ефективного критерію розгалуження, тобто вибір вузлів, атрибутів та характеристик дискретних об'єктів для схем ЛДК та алгоритмів АДК. Ці основні питання детально розглядаються в роботі [24], яка розглядає якісну оцінку та інформативність окремих дискретних ознак, їх наборів та фіксованих комбінацій, що в кінцевому підсумку забезпечує ефективну реалізацію механізму розгалуження в логічній/алгоритмічній структурі дерева. Проблеми, пов'язані з конвергенцією процесу побудови дерева класифікації, включаючи вибір критеріїв зупинки для синтезу логічних та алгоритмічних дерев, залишаються значними [25]. Концепція дерев класифікації передбачає використання не тільки окремих атрибутів і ознак об'єктів, але і їх комбінацій і наборів в якості ознак, атрибутів і вузлів деревовидної структури розпізнавання. Використовуючи незалежні індивідуальні (автономні) алгоритми розпізнавання (оцінені за допомогою навчальних даних) замість атрибутів об'єктів як вершин, реалізується нова структура АДК [21–24]. Дане націлене на вивчення структур АДК фіксованого типу в практичній площині.

Постановка завдання. Метою даного дослідження є побудова ефективних моделей АДК геологічних даних (ситуаційного стану) на основі концепції модифікованих структур АДК другого типу.

Виклад основного матеріалу.

Модифікований метод АДК другого типу. Наступна базова граф-схема синтезу дерева алгоритмів другого типу на основі розгалуженого вибору

узагальнених ознак дозволяю будувати структури АСТ довільної складності та ефективності (рис. 1).

Етап початкового відбору та оцінки незалежних алгоритмів класифікації. На початковому етапі необхідно відібрати та оцінити базовий набір (фіксовану множину) алгоритмів класифікації та розпізнавання (a_1, a_2, \dots, a_m) з початкової бібліотеки алгоритмів. Зауважимо, що ця процедура виконується на основі вибраного (зафіксованого) критерію ефективності з подальшим ранжуванням – інтерактивним або випадковим чином. Критерій ефективності може мінятися в залежності від типу структури ЛДК яка будується і не може бути змінена в процесі синтезу дерева класифікації. Набір автономних алгоритмів (a_1, a_2, \dots, a_m) , а також їх загальна кількість у наборі вибираються залежно від прикладних аспектів задачі та може бути підібрана навіть на основі повного перебору бібліотеки алгоритмів (зрозуміло, що з значними втратами апаратних ресурсів та процесорного часу). На початковому етапі синтезу моделі ЛДК другого типу, шляхом вибору (ранжування) набору алгоритмів класифікації та їх загальною кількістю можна керувати кінцевою структурною складністю дерева алгоритмів.

Етап синтезу структури дерева алгоритмів та узагальнених ознак. На наступному етапі стоїть центральне завдання – побудова повного регулярного дерева класифікації (фіксованої структури ЛДК), де відповідні яруси структури містять відібрані алгоритми класифікації (a_1, a_2, \dots, a_m) , зафіксовані на першому етапі побудови наборів класифікаторів.

Особливістю дерева алгоритмів другого типу є те, що в побудованій структурі дерева класифікації (ЛДК структурі) кожна вершина має два переходи на наступний рівень, що позначаються значенням з двійкового набору $\{0,1\}$. Саме тому структура дерева алгоритмів представлена за допомогою регулярної ЛДК конструкції. Виходячи з цього, всі атрибути (мітки) одного типу (алгоритми класифікації та згенеровані узагальнені ознаки) розташовані на кожному з рівнів цієї структури. У такій регулярній структурі дерева класифікації – вузли є незалежними алгоритмами (класифікаторами) (a_1, a_2, \dots, a_m) . Набори узагальнених ознак (УО) f_j також генеруються на етапі синтезу структури дерева алгоритму. Отже можна зробити висновок що дерево алгоритмів породжує дерево узагальнених ознак.

Ідея другого етапу синтезу структури дерева алгоритмів (моделі АДК II типу) полягає в процедурі синтезу набору узагальнених ознак f_j (вер-

шин дерева узагальнених ознак) на основі попередньо обраних наборів незалежних алгоритмів класифікації та розпізнавання a_i . Значимо, що загальна кількість УО f_j , які генерує відповідний алгоритм класифікації, залежить від вихідних параметрів моделі АДК та параметрів синтезу, специфіки прикладної задачі та ресурсних обмежень системи синтезу дерева класифікації.

В кінці другого етапу, після завершення формування набору синтезованих узагальнених ознак f_j для заданої прикладної задачі вони розташовуються у відповідних вузлах, ярусах дерева алгоритмів другого типу (будується структура дерева узагальнених ознак).

Центральним моментом даного етапу є вибір ефективних механізмів оптимального синтезу та розташування наборів УО в структурі моделі АДК. В залежності від типу АДК дана задача має багато варіантів розв'язку.

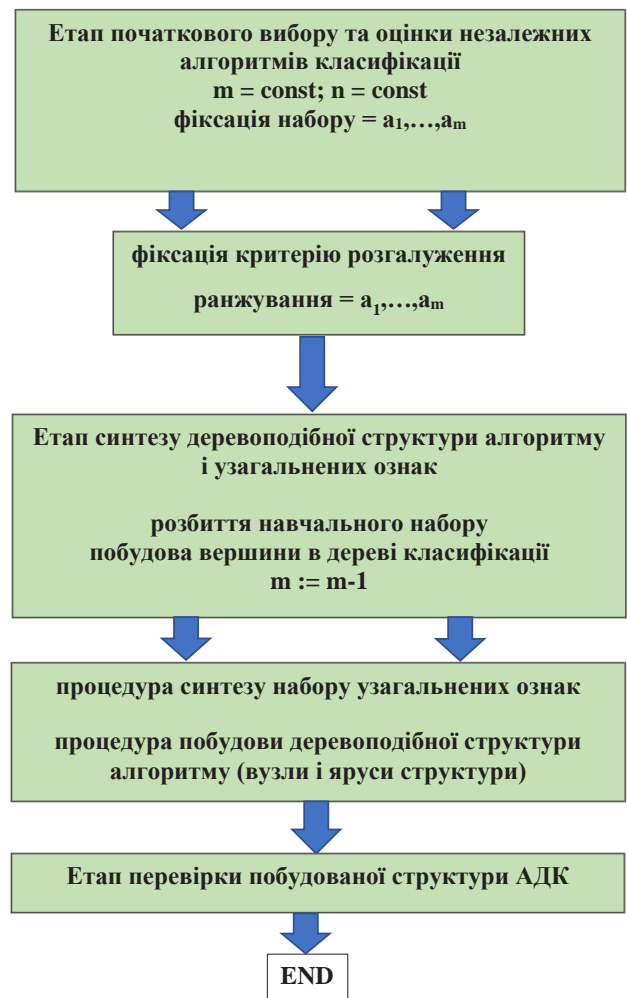


Рис. 1. Модифікований метод АДК другого типу

Етап фінальної оцінки ефективності побудованої структури АДК другого типу. На заключ-

ному етапі синтезу дерева алгоритмів другого типу необхідно перевірити побудовану модель АДК. Для кожного елемента (об'єкта) тестової вибірки обчислюються відповідні значення $\varphi(a_j)$. Обчислення цього значення здійснюється на основі набору раніше побудованих узагальнених ознак – для кожного вузла відповідного рівня дерева. Побудовані узагальнені ознаки визначають відповідний маршрут (обмежений класифікатор) в структурі дерева алгоритмів другого типу. Для такої структури кожен з вузлів дерева алгоритмів у разі можливої апроксимації об'єкта невідомої класифікації збільшує відповідний лічильник класу, що належить, і залишає його незмінним у разі помилки або збою класифікації. Дана процедура дозволяє провести остаточну оцінку ефективності побудованого дерева алгоритмів другого типу.

Експериментальна частина. Відмітимо, що запропонована схема побудови АДК дає змогу регулювати складність моделі дерева класифікації, що будується, будувати моделі з наперед заданою точністю, а сама структура дерева класифікації складається з різноманітних автономних алгоритмів класифікації як будівельних модулів (компонентів). Причому задача відбору моделі дерева класифікації серед набору побудованих АДК для конкретної задачі визначається набором параметрів, які мають визначальну важливість для поточної прикладної задачі (набору даних НВ).

Зрозуміло, що для порівняння та відбору конкретної моделі АДК з фіксованого набору, необхідно виділити найбільш важливі їх параметри (розмірність ознакового простору, кількість вершин, переходів, алгоритмів, тощо) та визначити їх похибку щодо масиву вхідних даних.

Принципово розглянути критерії якості отриманих інформаційних моделей, які залежать від похибки моделі, потужності початкового масиву даних НВ, об'єму тестової вибірки (кількості навчальних пар та розмірності ознакового простору задачі), кількості параметрів моделі тощо. Зрозуміло, що критично важливими параметрами побудованої моделі АДК, які необхідно мінімізувати, є помилки моделі відповідно на масивах даних навчальної та тестової вибірки та для кожного з класів, які задані початковою умовою поточної прикладної задачі.

Зауважимо, що принциповим моментом залишається питання зменшення складності структури АДК (йдеться про кількість ознак, алгоритмів у структурі АДК, загальну кількість вершин моделі АДК та загальну кількість переходів у структурі АДК), параметрів загальних

витрат пам'яті та процесорного часу інформаційної системи. Так, важливим показником якості побудованої моделі у вигляді дерева класифікації з урахуванням параметрів структури моделі ЛДК є загальний інтегральний показник якості в наступній формі:

$$Q_{Main} = \frac{Fr_{All}}{V_{All} \cdot \sum_i P_i} \cdot e^{-\frac{Er_{All}}{M_{All}}}. \quad (1)$$

Відмітимо, що в формулі (1) набір параметрів p_i представляє собою найбільш важливі характеристики побудованого дерева класифікації, що оцінюються:

1) Er_{All} – загальна кількість помилок моделі АДК на масивах даних початкових тестової та навчальної вибірки;

2) M_{All} – загальна потужність (об'єм) масивів даних навчальної та тестової вибірки;

3) Fr_{All} – кількість вершин отриманої моделі АДК із результативними значеннями f_r (функцій розпізнавання, тобто листів дерева класифікації);

4) V_{All} – представляє загальну кількість усіх типів вершин у структурі моделі АДК;

5) O_{Uz} – загальна кількість узагальнених ознак, що використовуються в моделі дерева класифікації;

6) P_{All} – загальна кількість переходів між вершинами в структурі побудованої моделі дерева класифікації;

7) N_{Alg} – загальна кількість різних автономних алгоритмів класифікації a_i , що використовуються в моделі дерева класифікації.

Відмітимо, що даний інтегральний показник якості моделі АДК буде приймати значення від нуля до одиниці. Чим меншим він буде, тим гіршою буде якість побудованого дерева класифікації, а чим більшим буде показник, тим кращою буде отримана модель.

На основі методів дерев класифікації, в Ужгородському національному університеті був розроблений програмний комплекс «Оріон» для генерації автономних систем розпізнавання. Алгоритмічна бібліотека системи нараховує 18 алгоритмів розпізнавання серед яких реалізовані схеми дерев алгоритмів трьох типів.

Базовою задачею на якій перевірялася ефективність методів дерев алгоритмів була задача розпізнавання геологічних даних – задача про розділення нафтоносних та водоносних пластів. Початкові параметри даної прикладної задачі класифікації геологічних даних представлені в (табл. 1).

В НВ представлена інформація про об'єкти двох класів. На етапі екзамени побудована система класифікації має забезпечити ефективне роз-

пізнання об'єктів невідомої класифікації відносно цих двох класів. Перед початком роботи навчальна вибірка була автоматично перевірена на коректність – пошук та видалення помилок першого роду. Відмітимо, що в системі реалізована схема донавчання та виправлення помилок в дереві класифікації (алгоритм ДВП).

Навчаюча вибірка представленої задачі складалася з 1342 об'єктів, з них нафтоносні – 761 об'єкти. Ефективність сконструйованої моделі АДК оцінювалася на тестовій вибірці об'єму 267 об'єктів. Дані навчаючих та тестових вибірок отримані на основі геологічної розвідки на території Закарпатської області в період з 2001 року по 2018 рік. Фрагмент основних результатів, приведених вище експериментів, побудованих моделей ЛДК/АДК різних типів представлений в (табл. 2).

В (Табл. 3) представлена інформація щодо часу генерації представлених моделей класифікації, загальної кількості вершин, елементарних та уза-

гальнених ознак на базовій апаратній конфігурації Intel i7-12700H. Всі побудовані схеми дерев класифікації (структур ЛДК/АДК) забезпечили необхідний рівень точності заданий умовою задачі, швидкодію та витрати робочої пам'яті системи.

Отже запропонований в роботі метод алгоритмічного дерева класифікації (методи АДК другого типу) порівнювалися з методом повного ЛДК та обмеженого методу селекції елементарних ознак та показав у цілому прийнятний результат.

Висновки. Побудовані моделі дерев класифікації (структури АДК/ЛДК) забезпечили необхідні якість та швидкість схем класифікації геологічних даних при достатньо компактній структурі самої конструкції дерева (параметр S_{main}). Набори незалежних алгоритмів класифікації, які були відібрані для генерації груп УО, також підтвердили свою ефективність в межах даної прикладної задачі. Найбільш якісною, з точки зору інтегральної оцінки, виявились саме моделі АДК різних

Таблиця 1

Початкові параметри задачі класифікації

Опис класів H_i задачі	Розмірність ознакового простору N	Загальна потужність початкової $NB - M$	Загальна кількість класів в $NB - l$	Співвідношення об'єктів різних класів в $NB - H_i/M$
Нафтоносні пласти (H_1)	(12/10)	1342	2	761/1342
Водоносні пласти (H_2)	(12/10)	1342	2	581/1342

Таблиця 2

Порівняльна таблиця побудованих моделей АДК/ЛДК для класифікації геологічних даних

№ моделі дерева класифікації	Метод синтезу структури дерева класифікації	Інтегральний показник якості моделі АДК Q_{Main}	Загальний показник структурної складності моделі дерева класифікації S_{Main}	Кількість помилок та відмов класифікації моделі ЛДК/АДК на масиві даних $E_{r_{All}}$
N1	Метод повного ЛДК на основі селекції елементарних ознак (розгалужений вибір ознак)	0,004786	121	7
N2	Метод ЛДК з одноразовою оцінкою важливості ознак	0,002271	144	12
N3	Обмежений метод побудови ЛДК	0,003193	97	16
N4	АДК (type I)	0,005287	52	10
N5	АДК (type II)	0,003033	64	8
N6	Обмежений метод побудови АДК	0,002654	55	14
N7	Дерево алгоритмів на основі гіперкуль	0,007221	31	6
N8	Дерево алгоритмів на основі гіперпаралелепіпедів	0,004418	54	19
N9	Дерево алгоритмів на основі гіпереліпсів	0,006476	30	8
N10	Дерево алгоритмів на основі гіперкубів	0,006251	37	11

Загальні структурні параметри побудованих моделей ЛДК/АДК

	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10
Загальний час синтезу дерева класифікації T_{All}	34 (с.)	21 (с.)	18 (с.)	65 (с.)	82 (с.)	55 (с.)	47 (с.)	56 (с.)	50 (с.)	98 (с.)
Кількість ярусів структури ЛДК/АДК R_{All}	12	10	9	26	30	23	21	24	22	34
Загальна кількість атрибутів / вершин структури ЛДК/АДК V_{All}	102	91	86	234	244	212	198	223	207	219
Загальна кількість елементарних / узагальнених ознак в структурі дерева класифікації O_{el}/O_{uz}	56 (ел.)	72 (ел.)	40 (ел.)	17 (уз.)	41 (уз.)	30 (уз.)	18 (уз.)	47 (уз.)	21 (уз.)	35 (уз.)

типів на основі базових геометричних класифікаторів. Причому синтезовані структури АДК видавали відносно невелику кількість помилок класифікації на даних навчальних та тестових вибірок. Модель повного АДК другого типу на основі геометричних класифікаторів показала близький результат ($Q_{main}=0,003033$) за рахунок наявності в схемі універсального алгоритму гіперсфер. Відмітимо, що структура АДК першого типу – то вона показала кращу якість ($Q_{main}=0,005287$) в порівнянні зі структурами дерев алгоритмів другого типу. Це пояснюється більш складною конструкцією моделі ($S_{main}=52$), як результат більшим часом генерації. Також треба зважати на особливості відібраних геометричних, які не завжди дозволяють проводити ефективну апроксимацію даних НВ. Недоліком представлених моделей АДК, який виявився в умовах даної задачі, є відносно великі часові затрати на етап синтезу самих моделей дерев класифікації в порівнянні зі структурами ЛДК. Різниця в порівнянні зі структурою ЛДК на основі розгалуженого вибору ознак з покроковою

оцінкою інформативності для моделей АДК першого типу склала майже 34 %.

Практична цінність отриманих результатів полягає в тому, що запропонований метод побудови моделей АДК (другого типу) дає можливість будувати економні та ефективні моделі класифікації заданої точності (даний метод був реалізований у бібліотеці алгоритмів системі “ОРІОН” для розв’язку різноманітних прикладних задач класифікації), які характеризуються великим ступенем універсальності відносно широкого кола прикладних задач. Відмітимо, що практичні застосування підтвердили працездатність побудованих моделей дерев класифікації та розробленого програмного забезпечення. В якості перспективи, майбутні дослідження можуть бути спрямовані в бік подальшого розвитку методів АДК (введення нових типів, схем дерев класифікації), оптимізації програмних реалізацій запропонованого методу АДК, а також його практичної апробації на множині реальних задач класифікації та розпізнавання.

Список літератури:

1. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, Berlin, Springer, 2008.
2. J.R. Quinlan, *Induction of Decision Trees*, *Machine Learning*, (1986) 81-10.
3. L.L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, *Classification and regression trees*, Boca Raton, Chapman and Hall/CRC, 1984.
4. M. Lupei, A. Mitsa, V. Repariuk, V. Sharkan, *Identification of authorship of Ukrainian-language texts of journalistic style using neural networks*, *Eastern-European Journal of Enterprise Technologies* 1 (2 (103)) (2020) 30-36. doi: 10.15587/1729-4061.2020.195041.
5. Y.V. Bodyanskiy, A.Y. Shafronenko, I.P. Pliss, *Credibilistic fuzzy clustering based on evolutionary method of crazy cats*, *System Research and Information Technologies*, 3/21 (2021) 110-119.

6. M. Miyakawa, *Criteria for selecting a variable in the construction of efficient decision trees*, IEEE Transactions on Computers 38(1) (1989) 130-141.
7. H. Koskimaki, I. Juutilainen, P. Laurinen, J. Roning, *Two-level clustering approach to training data instance selection: a case study for the steel industry*, in: Proceedings of the International Joint Conference on Neural Networks, IJCNN 2008, part of the IEEE World Congress on Computational Intelligence, IEEE, Los Alamitos. (2008) 3044-3049. doi: 10.1109/ijcnn.2008.4634228.
8. S.F. Jaman, M. Anshari, *Facebook as marketing tools for organizations: Knowledge management analysis*, in: Dynamic perspectives on globalization and sustainable business in Asia, IGI Global, Hershey. (2019) 92-105. doi: 10.4018/978-1-5225-7095-0.ch007.
9. V.E. Strilets, S.I. Shmatkov, M.L. Ugryumov et al, *Methods of machine learning in the problems of system analysis and decision making: monograph*. Karazin Kharkiv National University, (2020) 195.
10. R.L. De Mántaras, *A distance-based attribute selection measure for decision tree induction*, Machine learning 6(1) (1991) 81-92.
11. K. Karimi, H. Hamilton, *Generation and Interpretation of Temporal Decision Rules*, International Journal of Computer Information Systems and Industrial Management Applications 3 (2011) 314-323.
12. B. Kamiński, M. Jakubczyk, P. Szufel, *A framework for sensitivity analysis of decision trees*, Central European Journal of Operations Research 26 (1) (2017) 135-159.
13. H. Deng, G. Runger, E. Tuv, *Bias of importance measures for multi-valued attributes and solutions*, in: Proceedings of the 21st International Conference on Artificial Neural Networks, volume 2 of ICANN 2011, Springer-Verlag, Berlin. (2011) 293-300. doi: 0.1007/978-3-642-21738-8_38.
14. S.A. Subbotin, *Construction of decision trees for the case of low-information features*, Radio Electronics, Computer Science, Control 1 (2019) 121-130.
15. A. Shyshatskyi, *Complex Methods of Processing Different Data in Intellectual Systems for Decision Support System*. International Journal of Advanced Trends in Computer Science and Engineering, 9(4) (2020) 5583-5590. doi: <https://doi.org/10.30534/ijatcse/2020/206942020>
16. A. Painsky, S. Rosset, *Cross-validated variable selection in tree-based methods improves predictive performance*, IEEE Transactions on Pattern Analysis and Machine Intelligence 39(11) (2017) 2142-2153. doi:10.1109/tpami.2016.2636831.
17. D. Imamovic, E. Babovic, N. Bijedic, *Prediction of mortality in patients with cardiovascular disease using data mining methods*, in: Proceedings of the 19th International Symposium INFOTEH-JAHORINA, INFOTEH 2020, IEEE, Los Alamitos. (2020) 1-4.
18. S.B. Kotsiantis, *Supervised Machine Learning: A Review of Classification Techniques*, Informatica 31 (2007) 249-268.
19. Y.I. Zhuravlev, V.V. Nikiforov, *Recognition algorithms based on the calculation of estimates*, Cybernetics 3 (1971) 1-11.
20. I. Povkhan, O. Mules, O. Melnik, Y. Bilak, V. Polishcuk, *The Problem of Convergence of Classifiers Construction Procedure in the Schemes of Logical and Algorithmic Classification Trees*, CEUR Workshop Proceedings, Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2022), 3137 (2022) 1-13.
21. I. Povkhan, *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), 2608 (2020) 843-857.
22. Y.A. Vasilenko, E.Y. Vasilenko, I.F. Povkhan, *Conceptual basis of image recognition systems based on the branched feature selection method*, European Journal of Enterprise Technologies 7(1) (2004) 13-15.
23. I. Povkhan, M. Lupei, *The algorithmic classification trees*, in: Proceedings of the IEEE Third International Conference on Data Stream Mining & Processing, DSMP 2020, IEEE, Los Alamitos, (2020) 37-44.
24. I. Povkhan, M. Lupei, M. Kliap, V. Laver, *The issue of efficient generation of generalized features in algorithmic classification tree methods*, in: Proceedings of the International Conference on Data Stream Mining and Processing, DSMP 2020, Springer, Cham. (2020) 98-113.
25. I. Povkhan, *Classification models of flood-related events based on algorithmic trees*, Eastern-European Journal of Enterprise Technologies 6(4) (2020) 58-68. doi: <https://doi.org/10.15587/1729-4061.2020.219525>.
26. J. Rabcan, V. Levashenko, E. Zaitseva, M. Kvassay, S. Subbotin, *Application of Fuzzy Decision Tree for Signal Classification*, IEEE Transactions on Industrial Informatics 15(10) (2019) 5425-5434. doi: 10.1109/tii.2019.2904845.
27. P.E. Utgoff, *Incremental induction of decision trees*, Machine learning 4(2) (1989) 161-186. doi:10.1023/A:1022699900025.
28. L. Hyafil, R.L. Rivest, *Constructing optimal binary decision trees is NP-complete*, Information Processing Letters 5(1) (1976) 15-17.

29. H. Wang, M. Hong, *Online ad effectiveness evaluation with a two-stage method using a Gaussian filter and decision tree approach*, Electronic Commerce Research and Applications 1(35) (2019) Article 100852. doi: 10.1016/j.elerap.2019.100852.
30. I.L. Kaftannikov, A.V. Parasich, *Decision Tree's Features of Application in Classification Problems*, Bulletin of the South Ural State University, Ser. Computer Technologies, Automatic Control, Radio Electronics 15(3) (2015) 26-32. doi: 10.14529/ctcr150304.
31. I.F. Povhan, *Logical recognition tree construction on the basis a step-to-step elementary attribute selection*, Radio Electronics, Computer Science, Control. 2 (2020) 95-106.
32. S.L. Lin, H.W. Huang, *Improving Deep Learning for Forecasting Accuracy in Financial Data*, Discrete Dynamics in Nature and Society Volume 2020 (2020) Article ID 5803407. doi: 10.1155/2020/5803407.
33. E. Fincham, D. Gasevic, J. Jovanovic, A. Pardo, *From Study Tactics to Learning Strategies: An Analytical Method for Extracting Interpretable Representations*. IEEE Transactions on Learning Technologies, 12(1) (2020) 59–72. doi: <https://doi.org/10.1109/tlt.2018.2823317>
34. Ameisen E, *Building Machine Learning Powered Applications: Going from Idea to Product*, O'Reilly Media, California. (2020).
35. M. Sewak, *Deep Reinforcement Learning, Frontiers of Artificial Intelligence*, Springer, Berlin, 2020. doi: 10.1007/978-981-13-8285-7.
36. A. Kintonova, M. Mussaif, G. Gabdreshov, *Improvement of iris recognition technology for biometric identification of a person*, Eastern-European Journal of Enterprise Technologies, 6(2-120) (2023) 60-69.

Povkhan I.F., Saroz V.Ia., Leheza A.V. A MODIFIED METHOD OF THE SECOND-TYPE ALGORITHM TREE IN CLASSIFICATION TASKS

The article presents an effective method for synthesizing classification trees based on fixed initial information for the problem of recognizing geological data. The constructed algorithmic classification Tree of the second type will accurately classify the entire training sample, according to which the classification scheme is built, have high interpretability, simple structure and consist of autonomous classification algorithms, recognition as vertices of the graph scheme. The developed method of constructing algorithm trees allows you to work with training samples of a large volume of different types of discrete information, provides high accuracy of the scheme, efficiently uses the hardware resources of the system in the process of generating a final classification scheme, and allows you to build models with a pre-set accuracy. The paper offers a modular scheme for constructing classifiers in the form of classification tree structures and algorithm trees, which allows you to bypass the limitations of traditional decision tree methods. The modified method of the algorithm Tree of the second type can significantly reduce the computational complexity of the classification model in comparison with the algorithm trees of other structures and provide the necessary, configurable accuracy of the final classifier. An approach to synthesizing new recognition algorithms based on a library of already known algorithms and methods of recognition theory is proposed. This approach to the synthesis of classification models will allow us to comprehensively build new classifiers, classification models based on the modular principle. Based on the proposed method of the second type of algorithm tree, a set of models was built that provided an effective classification of the geological data array. The concept of algorithm trees of various types allows us to work effectively in classification problems with data from arbitrary scales, where information is set in the usual natural form for us. The constructed models of classification trees ensured that there were no errors on the data of the training and test sample, and confirmed the operability of the approach of trees of algorithms of the second type.

Key words: classification tree, algorithmic tree, discrete object, recognition algorithm, attribute.