

Повхан І.Ф.

Ужгородський національний університет

МОДЕЛІ ДЕРЕВ КЛАСИФІКАЦІЇ ПАВОДКОВИХ ЯВИЩ РІЧКИ УЖ ЗАКАРПАТСЬКОГО РЕГІОНУ

Робота присвячена виявленню простого й ефективного механізму, за допомогою якого можна будувати алгоритмічні дерева класифікації (моделі АДК) за фіксованою початковою інформацією у вигляді навчальної вибірки великого та надвеликого об'єму. Побудоване алгоритмічне дерево класифікації безпомилково класифікуватиме (розпізнаватиме) всю навчальну вибірку, за якою побудована модель, матиме мінімальну структуру (структурну складність) і складатиметься з компонентів – автономних алгоритмів класифікації та розпізнавання як вершин конструкції (атрибутів дерева).

Важливим моментом також є те, що концепція дерев класифікації не суперечить принциповій можливості як ознаки (вершини структури) дерева класифікації використовувати не тільки окремі атрибути (ознаки) об'єктів їх сполучення та набори, а й відбирати окремі незалежні алгоритми розпізнавання (оцінені за даними НВ), і на виході буде отримана нова структура – АДК. Запропонований метод побудови моделей АДК дозволяє працювати з навчальними вибірками великого об'єму різнотипної інформації та забезпечує високу швидкість та економність апаратних ресурсів у процесі генерації кінцевої схеми класифікації, будувати дерева класифікації з наперед заданою точністю. Пропонується синтез нових алгоритмів розпізнавання на основі бібліотеки (набору) вже відомих алгоритмів (схем) і методів. Тобто представлена ефективна алгоритмічна схема розпізнавання дискретних об'єктів на основі покрокової оцінки та вибору алгоритмів класифікації (узагальнених ознак) на кожному кроці синтезу схеми.

Одним із можливих напрямів застосування концепції алгоритмічних дерев класифікації є задачі, пов'язані із прогнозуванням і класифікацією паводкових явищ (на основі масивів метеорологічних і гідрологічних даних). Тому дослідження умов формування паводкових ситуацій у розрізі класифікації метеорологічних явищ за мінливих кліматичних умов є необхідним етапом для подальшого обґрунтування нових методик для розрахунків і прогнозів подолання кризових екологічних ситуацій. На базі запропонованої концепції алгоритмічних дерев класифікації побудована модель структури АДК, яка забезпечила класифікацію паводкових ситуацій для басейну річку Уж.

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

Постановка проблеми. 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. Нині не існує універсального підходу до їх розв'язання, запропоновано декілька досить загальних теорій і підходів, які дозволяють вирішувати багато типів (класів) задач, але їхнє прикладне застосування відрізняється великою чутливістю до специфіки самої задачі або предметної сфери застосування [1]. Багато теоретичних результатів отримано для спеціальних випадків і підзадач, причому слабким місцем вдалих реальних систем розпізнавання залишається необхідність виконання величезного об'єму обчислень та орієнтація на потужний апаратний інструментарій. Проте велика кількість прикладних задач у різних галузях природознавства, наприклад, у геології,

геофізиці, геохімії, медицині, соціології, археології, біології та ін., де вирішуються задачі класифікації з використанням програмних та апаратних систем, визначає інтенсивність та актуальність такого напряму досліджень [2–7]. Так, сьогодні відомо майже чотири тисячі методів та алгоритмів розпізнавання, які засновані на різноманітних підходах і концепціях, причому всі вони мають певні обмеження у використанні – точність, швидкодію, пам'ять. Важливо, що кожен із алгоритмів класифікації обмежений певною специфікою прикладних задач застосування (обмеження універсальності), що є найслабкішим місцем не тільки цих алгоритмів, але й інформаційних систем, які базуються на відповідних концепціях [6].

Нехай на деякій множині G дискретних об'єктів x задане розбиття R на скінчене число k підмножин (класів, образів) $H_i, (i=1, \dots, k)$, $G = \bigcup_{i=1}^k H_i$. Відповідні множини H_1, \dots, H_k будемо називати образами, а елементи множини

G – зображеннями або представниками образів H_1, \dots, H_k . Об'єкти (зображення) x задаються наборами значень деяких ознак $x_j, (j=1, \dots, n)$. Якщо $x \in H_i$, то будемо вважати, що цей об'єкт належить образу H_i . Загалом образи H_1, \dots, H_k можуть бути задані імовірнісними розподілами $p(H_1/x), \dots, p(H_k/x)$, де $p(H_i/x)$ – імовірність (або в неперервному випадку щільність імовірності) належності $x, (x \in H_i)$ образу H_i . Нехай умовою задачі задана деяка початкова НВ у вигляді послідовності навчальних пар такого вигляду:

$$(x_1, f_R(x_1)), \dots, (x_m, f_R(x_m)). \quad (1)$$

Причому крім початкової НВ задана також ТВ (тестова вибірка – набір об'єктів відомої класової належності) як деяка частина початкової НВ. Отже, за початковою умовою НВ – це сукупність (фіксована послідовність) деяких наборів (дискретних об'єктів), причому кожний набір – це сукупність значень деяких ознак (атрибутів) і значень деяких функцій розпізнавання (ФР) на цьому наборі. Тоді сукупність значень ознак – це деяке зображення (дискретний об'єкт), а значення функції (ФР) відносить це зображення до відповідного образу [3].

Отже, зазвичай стоїть загальна задача побудови моделі АДК із набором деяких параметрів p , структура L (дерева класифікації) якої була б оптимальною $F(L(p, x_i), f_R(x_i)) \rightarrow opt$ щодо початкових даних НВ (структури початкового масиву дискретних даних), причому нас цікавитиме побудова структур АДК для реальних навчальних вибірок метеорологічних даних.

Аналіз останніх досліджень і публікацій. Наше дослідження продовжує цикл робіт, присвячених проблематиці деревоподібних схем розпізнавання (класифікації) дискретних об'єктів [2–5; 8–10]. У них піднімаються питання побудови, використання й оптимізації логічних дерев. Так, із [3] відомо, що результуюче правило класифікації (схема), яке побудоване довільним методом або алгоритмом розгалуженого вибору ознак, має деревоподібну логічну структуру. Логічне дерево складається з вершин (ознак), що групуються за ярусами й отримані на певному кроці (етапі) побудови дерева розпізнавання [6]. Важливою задачею, яка виникає з [8], є задача синтезу дерев розпізнавання, котрі будуть представлятися фактично деревом (графом) алгоритмів. На відміну від наявних методів, головною особливістю деревоподібних систем розпізнавання є те, що важливість окремих ознак (групи ознак чи алгоритмів) визначається відносно функції, яка задає розбиття

об'єктів на класи [12]. Так, у роботі [15] піднімаються принципові питання генерації дерев рішень для випадку малоінформативних ознак. Здатність структур ЛДК/АДК виконувати одномірне розгалуження для аналізу впливу (важливості, якості) окремих змінних дає можливість працювати зі змінними різних типів у вигляді предикатів (у випадку АДК – відповідними автономними алгоритмами класифікації та розпізнавання) [14]. Така концепція логічних дерев активно використовується в інтелектуальному аналізі даних, де кінцева мета полягає в синтезі моделі, яка прогнозує значення цільової змінної на основі набору початкових даних на вході системи [15].

Важливим моментом також є те, що концепція дерев класифікації не суперечить принциповій можливості як ознаки (вершини структури) дерева класифікації використовувати не тільки окремі атрибути (ознаки) об'єктів їх сполучення та набори, а й відбирати окремі незалежні алгоритми розпізнавання (оцінені за даними НВ). На виході буде отримана нова структура – АДК [14]. Саме структурам у практичній площині і буде присвячена наша робота.

Постановка завдання. Отже, з огляду на все вищесказане метою роботи буде отримання методу побудови структур АДК для класифікації паводкових явищ річки Уж (на основі оцінки та ранжування набору автономних алгоритмів розпізнавання та класифікації для генерації структури дерева класифікації, причому на кожному кроці розгалуження дерева класифікації апроксимується певна частина НВ (або її підмножина)).

Виклад основного матеріалу дослідження. За аналогією з методом апроксимації НВ набором оцінених елементарних ознак, поданим у роботі [3], представимо головну ідею методів алгоритмічних дерев класифікації, яка полягає в апроксимації масиву початкової НВ набором автономних різнотипних алгоритмів класифікації.

Загальна концепція методів АДК

Нехай на початку задана початкова НВ загального вигляду (1) як послідовність навчальних пар відомої класифікації (потужності m) та деяка система (набір) незалежних та автономних алгоритмів розпізнавання (класифікації) для початкової НВ $a_1(x), a_2(x), \dots, a_m(x)$. Введемо такі множини, які представляють розбиття даних НВ відповідними алгоритмами класифікації a_i :

$$G_{a_1, \dots, a_i} = \{x \in G / a_i(x) = 1\}, (i=1, \dots, n). \quad (2)$$

Для спрощення пояснень кожний з автономних алгоритмів класифікації $a_i(x)$ генерує на виході

значення лише в межах бінарної множини $\{0,1\}$, тобто тут $a_i(x)=1$ у разі вдалої класифікації об'єкта x та $a_i(x)=0$ у протилежному випадку.

Система множин G_{a_1, \dots, a_i} фактично представлятиме повне поетапне розбиття множини G (зі зростання величини i – тобто задіяних алгоритмів класифікації), яке реалізується незалежними алгоритмами a_1, a_2, \dots, a_M . Залежно від початкового обрання набору алгоритмів класифікації a_1, a_2, \dots, a_M деякі з множин G_{a_1, \dots, a_i} можуть бути пустими (випадок непридатності одного конкретного або декількох алгоритмів для апроксимації поточної НВ).

Через величину S_{a_1, \dots, a_i} позначимо кількість входжень у початкову НВ тих пар навчальних $(x_s, f_R(x_s))$, $(1 \leq s \leq m)$, які задовольняють базову умову належності $x_s \in G_{a_1, \dots, a_i}$. Відповідно, через величину S_{a_1, \dots, a_i}^j , $j=0, 1, \dots, k-1$ позначимо кількість входжень у НВ тих пар $(x_s, f_R(x_s))$, $(s=1, 2, \dots, m)$, які задовольняють умови $x_s \in G_{a_1, \dots, a_i}$ та $f_R(x_s)=j$. Отже, з огляду на вищесказане та за аналогією до методів селекції наборів елементарних ознак [12] можна ввести такі величини, які доцільно розглядати як певний критерій розгалуження у структурі АДК:

$$\delta_{a_1, \dots, a_i} = \frac{S_{a_1, \dots, a_i}}{m}, \psi_{a_1, \dots, a_i}^j = \frac{S_{a_1, \dots, a_i}^j}{S_{a_1, \dots, a_i}}, \rho_{a_1, \dots, a_i} = \max_j \psi_{a_1, \dots, a_i}^j. \quad (3)$$

Зауважимо, що, якщо об'єкт $x_s \notin G_{a_1, \dots, a_i}$ для всіх $s=1, \dots, m$, тоді зрозуміло, що δ_{a_1, \dots, a_i} та $\psi_{a_1, \dots, a_i}^j = 0$, при $(j=0, 1, \dots, k-1)$. Величина δ_{a_1, \dots, a_i} характеризує частоту входжень членів послідовності x_1, x_2, \dots, x_m (дискретних об'єктів) у множину G_{a_1, \dots, a_i} , а відповідно, величина ψ_{a_1, \dots, a_i}^j характеризує частоту належності об'єкту деякого x образу (класу) H_j за умови, що $x \in G_{a_1, \dots, a_i}$. Тут слід зауважити, що 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 $a_y(x)=1$. Тоді величина δ_{a_1, \dots, a_i} характеризує інформаційну ефективність розпізнавання належності деякого об'єкта x до одного з класів H_0, H_1, \dots, H_{k-1} звичайно, за умови, що $x \in G_{a_1, \dots, a_i}$. На наступному етапі знову виникає принципове питання щодо належності об'єкта x класам H_0, H_1, \dots, H_{k-1} (питання формування правила класифікації). Зрозуміло, що тут також слід віднести об'єкт x до того класу H_j , для якого виконується просте співвідношення:

$$\rho_{a_1, \dots, a_i} = \psi_{a_1, \dots, a_i}^j. \quad (4)$$

Зауважимо, що тут $\{0 \leq j \leq k-1\}$, а співвідношення (4) представляє деяке правило класифікації, причому зрозуміло, що чим більше величина ρ_{a_1, \dots, a_i} , тим вища його ефективність. Оскільки

єдиною інформацією, яка представляє розбиття образів H_0, H_1, \dots, H_{k-1} , є початкова НВ, під класом H_j розуміється сукупність всіх навчальних пар $(x_s, f_R(x_s))$ НВ, які задовольняють співвідношення $f_R(x_s)=j$, тобто умову належності. Аналогічно середня ефективність розпізнавання набору образів H_0, H_1, \dots, H_{k-1} , що задані даними НВ за допомогою алгоритмів розпізнавання $\alpha_1, \alpha_2, \dots, \alpha_n$ оцінюється такою величиною:

$$F_S(\alpha_1, \alpha_2, \dots, \alpha_n) = \sum_{a_1, \dots, a_i} \delta_{a_1, \dots, a_i} * \rho_{a_1, \dots, a_i}. \quad (5)$$

Отже, величину $F_S(\alpha_1, \dots, \alpha_n)$ можна вважати оцінкою апроксимації початкової НВ за допомогою набору незалежних алгоритмів класифікації $\alpha_1, \alpha_2, \dots, \alpha_n$. З огляду на саму ідею алгоритмічного дерева класифікації представлену вище величину $F_S(\alpha_1, \dots, \alpha_n)$ можна отримати ще з таких міркувань – функцією розпізнавання F назвемо деяке відображення, яке кожному набору a_1, \dots, a_i ставить у відповідність деякий елемент множини $\{0, 1, \dots, k-1\}$ (тобто відповідний номер класу). Таким чином, ФР F представляє функцію вигляду $F(a_1, \dots, a_i)$, де a_1, \dots, a_i приймають значення із множини $\{0, 1\}$. Відповідно, ФР $F(a_1, \dots, a_i)$ об'єкт x , $(x \in G)$ однозначно належить до того класу (образу) H_j , $(0 \leq j \leq k-1)$, для якого виконується співвідношення:

$$F(a_1, \dots, a_i) = l, (0 \leq l \leq k-1). \quad (6)$$

Експериментальна частина

Одним із можливих напрямів застосування концепції алгоритмічних дерев класифікації є задачі, пов'язані із прогнозуванням і класифікацією паводкових явищ (на основі масивів метеорологічних і гідрологічних даних). Відомо, що паводкові явища завдають величезної шкоди як економічному, так і екологічному сектору народного господарства: затоплюють і виводять із господарського обороту сільськогосподарські угіддя, підтоплюють населені пункти, руйнують житлові будинки та промислові підприємства, греблі, шосе та залізниці, трапляються і людські жертви, що надзвичайно актуально для Закарпатського регіону. Тому дослідження умов формування паводкових ситуацій у розрізі класифікації метеорологічних явищ за мінливих кліматичних умов є необхідним етапом для подальшого обґрунтування нових методик для розрахунків і прогнозів подолання кризових екологічних ситуацій. Відзначимо, що режим поверхневого водного стоку басейну річки Уж характеризується значною територіальною нерівномірністю, зумовленою кліма-

тичними, температурними факторами та впливом гірських масивів Карпат, які мають визначальний вплив для формування сезонних стоків. Для аналізу умов і причин формування паводкових явищ на річці Уж Закарпатської області в осінньо-весняний період використовувалися дані двох гідрологічних (спостережних) постів із періодом моніторингу від початку 1992 і по 2010 р. включно (на ділянці спостереження 30 км). За водним режимом річку Уж можна віднести до річок із паводковим режимом Карпатського підтипу, де паводкові ситуації спостерігаються переважно в осінньо-весняний період. Зазвичай такі паводкові ситуації називають паводками холодного періоду, причому вони спостерігаються в період із жовтня-листопада по квітень-травень. Зрозуміло, що з поправкою на сучасні кліматичні зміни межі теплого і холодного періодів не є постійними з року в рік, в основу визначення їхнього початку і кінця були покладені базові метеорологічні умови (температура повітря, види атмосферних опадів і стан снігового покриву). Початкові параметри цієї прикладної задачі класифікації паводкових явищ річки Уж представлені в табл. 1. Загальна 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 в Ужгородському національному університеті були побудовані моделі класифікації паводкових явищ для річки Уж за період 1992–2010 у вигляді структур 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 і схем розпізнавання). Так, НВ складалася з об'єктів (кожен із яких описується 18 ознаками) трьох базових класів, а на етапі екзамени побудована система класифікації (модель АДК) має забезпечити ефективне розпізнавання об'єктів невідомої класифікації щодо цих трьох класів. Загальні параметри (характеристики) побудованих структур (моделей АДК) представлені в табл. 2.

На початковому етапі роботи програмної системи навчальна вибірка була автоматично перевірена на коректність (пошук і видалення однакових об'єктів різної належності – помилки першого роду). Зауважимо, що в масиві початкової навчальної інформації переважали навчальні пари класу H_3 (об'єкти ситуаційного стану нейтральної зони, зеленого маркеру), на другому місці зі значним відривом за кількістю знаходилися навчальні пари класу H_2 (об'єкти ситуаційного стану спостережної зони, жовтого маркеру), на третьому – 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 здебільшого переходить у кризовий стан (червону зону паводкового явища) класу H_1 . Масив НВ складався із 8 391 об'єкта (набору 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 із постів спостереження, причому масив ТВ становив відокремлену частину початкової НВ (складався з дискретних об'єктів відомої класифікації).

Важливим моментом є аналіз критеріїв якості отриманих інформаційних моделей, які залежать від похибки моделі, потужності початкового масиву даних НВ та ТВ (кількість навчальних пар і розмірність ознакового простору задачі), кількості структурних параметрів моделі і т. д. Зрозуміло, що критично важливими параметрами побудованої моделі АДК, котрі необхідно мінімізувати, є помилки моделі на масивах даних НВ, ТВ та для кожного із класів (частин, підмножин початкової НВ), які задані початковою умовою поточної прикладної задачі. Одним із найважливіших показників, що характеризує базові властивості отриманих моделей АДК, є базовий показник узагальнення даних початкової НВ деревом класифікації (моделлю класифікації), який розраховується таким чином:

Таблиця 1

Початкові параметри задачі класифікації паводкових явищ басейну річки Уж

Номер поста спостереження, №	Розмірність ознакового простору – N	Потужність масиву даних початкової НВ – M	Потужність масиву даних ТВ – S	Загальна кількість класів за розбиттям даних НВ – l	Відношення об'єктів різних класів НВ ($H_1 / H_2 / H_3$)
1	18	4252	500	3	73/102/4107
2	18	4139	500	3	68/97/3974

$$I_{Main} = \frac{m * O_{Uz}}{Fr_{All} + V_{All} + N_{All} + 2P_{All}} \quad (9)$$

Цей показник узагальнення моделі дерева класифікації (структури АДК) відображає його базові параметри (характеристики) дерев класифікації та може бути застосований як критерій оптимальності у процедурі оцінки довільної деревоподібної схеми розпізнавання. Для довільної прикладної задачі важливо максимізувати параметр I_{Main} (показник узагальнення моделі АДК), що дозволяє добитися найбільш оптимальної структури дерева (моделі) класифікації та забезпечує фактично максимальний стиск даних початкової НВ (представити масив початкових даних мінімальним за структурною складністю деревом) [28]. Важливим показником якості побудованої моделі у вигляді дерева класифікації з урахуванням параметрів структури моделі АДК є загальний інтегральний показник якості, представлений у формі:

$$Q_{Main} = \frac{Fr_{All}}{O_{Uz} * \sum p_i} * e^{-\frac{Er_{All}}{M_{All}}} \quad (10)$$

У цих функціоналах (9–10) параметри мають таку інтерпретацію:

Er_{All} – загальна кількість помилок моделі АДК на масивах даних початкових тестової та навчальної вибірки – $Er_{All} = En_{tr} + Et_{tr}$;

M_{All} – загальна потужність цих двох масивів даних – $M_{All} = m + T$, де m та T відповідні потужності НВ та ТВ;

Fr_{All} – характеризує загальну кількість вершин отриманої моделі АДК із результируючими значеннями f_R (ФР, тобто листів дерева класифікації);

O_{Uz} – представляє загальну кількість всіх узагальнених ознак (наборів УО) у структурі моделі АДК;

V_{All} – представляє загальну кількість всіх типів вершин (крім результируючих ФР) у структурі моделі АДК;

N_{All} – загальна кількість різних автономних алгоритмів класифікації, що використовуються в моделі дерева класифікації;

P_{All} – загальна кількість переходів між вершинами (ярусами) у структурі побудованої моделі дерева класифікації.

Набір параметрів p_i для інтегральної оцінки якості моделі дерева класифікації становить найбільш важливі характеристики дерева класифікації (відповідно до структур ЛДК/АДК), що оцінюється (наприклад, кількість елементарних ознак або узагальнених ознак, які використовуються в моделі дерева класифікації, кількість переходів між вершинами, ярусами дерева класифікації тощо).

Висновки. Побудовані дерева класифікації (моделі АДК) забезпечили необхідні якість і швидкість схем класифікації паводкових явищ

Таблиця 2

Порівняльна таблиця моделей дерев 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

річки Уж за достатньо компактною структурою самої конструкції дерева. Набори незалежних алгоритмів класифікації, які були відібрані для генерації груп УО, також підтвердили свою ефективність у межах цієї прикладної задачі. Можливим шляхом подальших досліджень може бути розширення переліку алгоритмів класифікації у схемі АДК, а також додаткові умови й обмеження щодо генерації наборів УО для кожного кроку схеми дерева класифікації (структури АДК). Представлені моделі дерев класифікації можна застосовувати для оцінки загального стану басейну річки Уж (на ділянці спостереження) та виявлення ситуації червоної (паводкової) зони на основі поточних замірів постів спостережень.

Відзначимо, що проведені практичні випробування структур (моделей) АДК підтвердили працездатність математичного забезпечення та запропонованих методів та алгоритмів побудови АДК, розробленого програмного забезпечення, що дозволяє зробити рекомендацію щодо використання цього підходу (концепції моделей АДК) та його програмної реалізації для широкого спектру прикладних задач класифікації та розпізнавання у практичній площині. Єдиним прин-

циповим моментом, на який треба зважати при генерації моделей АДК, є те, що зазвичай витрати робочої пам'яті та процесорного часу інформаційної системи є значно більшими порівняно зі структурами (моделями) ЛДК, причому значною мірою це залежить від особливостей реалізації алгоритмів розпізнавання (класифікаторів), кількості алгоритмів у схемі АДК, схеми (типу моделі) структури АДК, яка генерується.

Отже, з огляду на все вищесказане перспективним напрямом розв'язку задачі класифікації ситуаційного стану річки Уж може бути також застосування методів, алгоритмів і схем побудови моделей ЛДК із можливістю їх порівняння (ефективності, структурної складності та швидкодії класифікації) з побудованими структурами (моделями) АДК, розширення об'єму бібліотеки алгоритмів розпізнавання (класифікаторів), оптимізації програмних реалізацій запропонованих у цьому дослідженні методів побудови АДК (за взірць такої ефективної програмної оптимізації можна взяти еволюційний перехід від алгоритму C4.5 до C5.0), а також практичної апробації побудованих моделей дерев класифікації на множині даних НВ більшої кількості постів спостереження басейну річки Уж.

Список літератури:

1. Zheng Z., Kohavi R., Mason L. Real world performance of association rule algorithms. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* / ed. by F. Provost, R. Srikant. 2001. P. 401–406.
2. Василенко Ю.А., Василенко Е.Ю., Ващук Ф.Г., Повхан І.Ф. Концептуальна основа систем розпізнавання образів на основі метода розгалуженого вибору ознак. *European Journal of Enterprise Technologies*. 2004. № 7 [1]. С. 13–15.
3. Василенко Ю.А., Василенко Е.Ю., Ващук Ф.Г., Повхан І.Ф. Метод розгалуженого вибору ознак в математичному конструюванні багаторівневих систем розпізнавання образів. *Штучний Інтелект*. 2003. № 7. С. 246–249.
4. Василенко Ю.А., Василенко Е.Ю., Ващук Ф.Г., Повхан І.Ф. Проблема оцінки складності логічних дерев розпізнавання та загальний метод їх оптимізації. *European Journal of Enterprise Technologies*. 2011. № 6 / 4 (54). С. 24–28.
5. Повхан І.Ф., Ващук Ф.Г. Загальна оцінка мінімізації деревоподібних логічних структур. *European Journal of Enterprise Technologies*. 2012. № 1 / 4 (55). С. 29–33.
6. Quinlan J.R. Induction of Decision Trees. *Machine Learning*. 2008. № 1. P. 1–81.
7. Vtoghoff P.E. Incremental Induction of Decision Trees. *Machine Learning*. 2009. № 4. P. 161–186.
8. Повхан І.Ф. Проблема функціональної оцінки навчальної вибірки в задачах розпізнавання дискретних об'єктів. *Вчені записки Таврійського національного університету. Серія: технічні науки*. 2018. Т. 29 (68). № 6. С. 217–222.
9. Povhan I. General scheme for constructing the most complex logical tree of classification in pattern recognition discrete objects. *Електроніка та інформаційні технології*. 2019. Вип. 11. С. 112–117.
10. Povhan I. Generation of elementary signs in the general scheme of the recognition system based on the logical tree. *Електроніка та інформаційні технології*. 2019. Вип. 12 С. 20–29.
11. Srikant R., Agrawal R. Mining generalized association rules. *Future Generation Computer Systems*. 1997. Vol. 13. № 2. P. 161–180.
12. Povhan I. Designing of recognition system of discrete objects. *2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine*. Lviv. P. 226–231.
13. Mitchell T. *Machine learning*. New York : McGraw-Hill, 1997. 432 p.

14. Povkhan I.F. Features of synthesis of generalized features in the construction of recognition systems using the logical tree method. *Information technologies and computer modeling ITKM-2019* : materials of the international scientific and practical conference, Ivano-Frankivsk, May 20–25, 2019. Ivano-Frankivsk, 2019. P. 169–174.
15. Subbotin S.A. Methods of sampling based on exhaustive and evolutionary search. *Automatic Control and Computer Sciences*. 2013. Vol. 47. № 3. P. 113–121.
16. Povhan I. Question of the optimality criterion of a regular logical tree based on the concept of similarity. *Electronics and information technology*. 2020. Vol. 13. P. 12–16.
17. Miyakawa M. Criteria for selecting a variable in the construction of efficient decision trees. *IEEE Transactions on Computers*. 1989. Vol. 38. № 1. P. 130–141.

Povkhan I.F. MODELS OF TREES FOR CLASSIFICATION OF FLOOD EVENTS OF THE UZH RIVER IN THE TRANSCARPATHIAN REGION

The paper is devoted to the identification of a simple and effective mechanism that can be used to build algorithmic classification trees (ACT models) based on fixed initial information in the form of a large and ultra-large sample. An algorithmic classification tree is constructed. it will accurately classify (recognize) the entire training sample that the model is based on, have a minimal structure (structural complexity), and consist of components of Autonomous classification and recognition algorithms as construction vertices (tree attributes).

The important point is that the concept of classification trees is not contrary to the principle of signs (vertex structure) of a classification tree to use not only individual attributes (characteristics) of objects, their combinations and sets, but if you go ahead and not be considered as a branching object attributes (characteristics) and select independent algorithms (measured according to TS) then the output will be a new structure of ACT. The proposed method for constructing ACT models allows you to work with training samples of a large amount of different types of information and provides high speed and economy of hardware resources in the process of generating the final classification scheme, build classification trees with a predetermined accuracy. We propose an approach for synthesizing new recognition algorithms based on a library (set) of already known algorithms (schemes) and methods. In other words, we present an effective algorithmic scheme for recognizing discrete objects based on step-by-step evaluation and selection of classification algorithms (generalized features) at each step of the scheme synthesis.

One of the possible areas of application of the concept of algorithmic classification trees is problems related to the prediction and classification of flood events (based on arrays of meteorological and hydrological data). Therefore, the study of conditions for the formation of flood situations in the context of classification of meteorological phenomena under changing climatic conditions is a necessary stage for further substantiation of new methods for calculations and forecasts of overcoming environmental crisis situations. Based on the proposed concept of the classification tree algorithm, a model of the ACT structure was constructed, which provided classification of flood situations for the Uzh river basin.

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