

**I. Ф. Повхан***Ужгородський національний університет, м. Ужгород, Україна***МЕТОД ОБМЕЖЕНИХ СТРУКТУР ЛОГІЧНИХ ДЕРЕВ У ЗАДАЧІ КЛАСИФІКАЦІЇ ДИСКРЕТНИХ ОБ'ЄКТІВ**

Розглядається проблема побудови моделі логічних дерев класифікації на підставі обмеженого методу селекції елементарних ознак для масивів геологічних даних. Запропоновано метод апроксимації масиву реальних даних набором елементарних ознак з фіксованим критерієм зупинки процедури розгалуження на етапі побудови дерева класифікації. Даний підхід дає змогу забезпечити необхідну точність моделі, знизити її структурну складність та досягти необхідних показників ефективності. Розроблено обмежений метод побудови дерев класифікації, який спрямований на побудову тільки тих шляхів (ярусів) структури дерева класифікації, де є найбільша кількість помилок (усіх типів) класифікації. Такий підхід синтезу моделі розпізнавання дає можливість досить ефективно регулювати складність (точність) моделі дерева класифікації, що будеться, причому 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 навчальної вибірки. Обмежена схема синтезу дерев класифікації дає змогу будувати моделі майже на 20 % швидше. Побудоване логічне дерево класифікації буде безпомилково класифікувати (розпізнавати) всю навчальну вибірку за якою побудована модель, мати мінімальну структуру (структурну складність) та складатися з компонентів – наборів елементарних ознак як вершини конструкції, атрибутів дерева. На підставі запропонованої модифікації методу селекції елементарних ознак розроблено програмне забезпечення, яке дає змогу роботу з набором різнотипних прикладних задач. Пропонується підхід синтезу нових моделей розпізнавання на підставі обмеженої схеми логічних дерев та вибору параметрів препрунінгу. Тобто розроблена ефективна схема розпізнавання дискретних об'єктів на підставі покрокової оцінки і вибору наборів атрибутів (узагальнених ознак) за відібраними шляхами в структурі дерева класифікації на кожному кроці синтезу схеми.

**Ключові слова:** логічне дерево класифікації; розпізнавання образів; класифікація; ознака; критерій розгалуження.

**Вступ**

Важливим сегментом області застосувань концепції дерев рішень залишають методи дерев класифікації, регресійні дерева, які активно використовуються як для задач теорії штучного інтелекту, засобом підтримки прийняття рішень, аналізу великих масивів даних так і в суміжних практичних галузях економіки, управління тощо [22]. Основні наявні методи оброблення навчальних вибірок при побудові функції розпізнавання не дають змогу досягнути потрібного рівня точності системи розпізнавання та регулювати їх складність у процесі конструювання даних систем [16], [34]. Цей недолік відсутній у методах побудови систем розпізнавання, які базуються на методах логічних дерев класифікації.

Дерева класифікації є одним з базових методів автоматичного аналізу даних. Так перші всебічні дослідження та концепції ідеї застосування дерев рішень (класифікації) ведуть початок від робіт С. Ховленда і Е. Ханта [6]. Відзначимо, що важливою особливістю дерев класифікації є гнучкість, тобто здатність логічних дерев класифікації (ЛДК) послідовно враховувати та досліджувати ефект впливу окремих змінних, атрибутів структури. Відповідно, є ще цілий ряд причин, що забезпечують структурам ЛДК більшу гнучкість, ніж традиційні методи та інструменти аналізу даних. Так, здатність ЛДК виконувати одномірне розгалуження для аналізу впливу (важливості, якості) окремих змінних дає можливість працювати зі змінними різних типів у вигляді предикатів [23].

Базовим питанням в схемі побудови дерева класифікації залишається проблема відбору правила зупинки розгалуження (критерій зупинки побудови ЛДК). Аналізуючи наявні методи та алгоритми побудови логічних дерев, можна виділити три наступні базові підходи в цьому напрямку – при побудові ЛДК використовувати статистичні методи оцінки потреби подальшого розбиття (розгалуження) множини (це так звана рання зупинка – *prepruning*), при побудові ЛДК використовують схему обмеження глибини (кількості ярусів) логічного дерева.

Так звернемо увагу, що в схемі методів побудови ЛДК, наприклад в роботах [20], [33] є принциповий недолік, який пов'язаний з тим що зі зростанням кількості вершин в структурі логічного дерева (збільшення ярусів ЛДК) кількість елементарних ознак в дереві значно збільшується. Звичайно таке ускладнення результуючого ЛДК негативно впливає на апаратні можливості системи розпізнавання (пам'ять, процесорний час). Для того, щоби подолати ці негативні моменти буде запропонована модифікація (обмежений метод) методу ЛДК на підставі поетапної селекції елементарних ознак.

В практичній площині досить часто алгоритми та методи побудови ЛДК на виході дають структурно складні логічні дерева (в плані кількості вершин, кількості розгалужень, належності до класу нерегулярних дерев), які нерівномірно заповнені даними, мають різну кількість розгалужень. Такі складні деревоподібні структури досить складно сприймаються для зовнішнього аналізу за рахунок великої кількості вузлів (вершин) та

великої кількості покровових розбиттів початкової навчальної вибірки (НВ), як містять мінімальну кількість об'єктів (можливо навіть одиничні об'єкти в найгіршому випадку). Зрозуміло, що значно краще мати деяке ЛДК з мінімальною кількістю вершин (вузлів), яким би відповідала абсолютна більшість об'єктів початкової НВ. Саме на цьому етапі виникає принципове питання теорії ЛДК – питання можливої побудови всіх варіантів логічних дерев, які відповідають початковій НВ та відбору мінімального за глибиною (кількістю ярусів) логічного дерева [15], [17], [29], [30], [31].

*Об'єкт дослідження* – процеси синтезу логічних дерев класифікації різних типів та схем.

*Предмет дослідження* – методи та алгоритми побудови логічних дерев класифікації (дерев рішень).

*Мета роботи* – створення ефективного та універсального методу побудови моделей класифікації на підставі концепції ЛДК для масивів дискретної інформації на підставі процедури препрунгу.

Для досягнення зазначеної мети визначено такі основні завдання дослідження: провести аналіз концепції синтезу структур ЛДК; побудова загального методу синтезу обмежених структур ЛДК на підставі процедури препрунгу.

*Наукова новизна отриманих результатів дослідження* – вперше розроблено обмежений метод побудови ЛДК на підставі селекції елементарних ознак з постійною оцінкою їх важливості на кожному кроці генерації дерева класифікації з можливістю пріоритетної побудови фіксованих блоків конструкції ЛДК.

*Практична значущість результатів дослідження* – запропонований обмежений метод побудови моделей ЛДК дає можливість будувати економічні та ефективні моделі класифікації заданої точності (даний метод був реалізований в бібліотеці алгоритмів системи "ОРИОН III" для розв'язку різноманітних прикладних задач класифікації). Причому проведені практичні застосування підтвердили працездатність побудованих моделей ЛДК та розробленого програмного забезпечення. Як перспектива, подальші дослідження можна спрямувати в бік подальшого розвитку методів ЛДК, оптимізації програмних реалізацій запропонованого обмеженого методу побудови ЛДК а також його практичної апробації на множині реальних задач класифікації та розпізнавання.

*Матеріали та методи дослідження.* Нехай задана НВ в такому вигляді:

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

Зауважимо, що тут  $x_i \in G, f_R(x_i) \in \{0, k-1\}, (i = \overline{1, m})$ ,  $m$  – кількість об'єктів з НВ,  $f_R(x_i)$  – деяка скінченнозначна функція, що задає розбиття  $R$  множини  $G$  на класи (образи)  $H_0, H_1, \dots, H_{k-1}$ . Відношення  $f_R(x_i) = l$  ( $l = \overline{1, k-1}$ ) означає  $x_i \in H_l, x_i = \{x_{i_1}, x_{i_2}, \dots, x_{i_n}\}$ ,  $x_{i_j}$  – значення  $j$ -ої ознаки для об'єкта  $x_i, (j = \overline{1, 2, \dots, n})$ ,  $n$  – кількість ознак в НВ.

Отже, НВ – це сукупність (точніше послідовність) деяких наборів, причому кожний набір – це сукупність значень деяких ознак та значень деяких функцій на цьому наборі. Можна ще сказати, що сукупність значень ознак – це деяке зображення, а значення функцій відносить це зображення до відповідного образу [19]. Ставиться задача побудови конструкції ЛДК  $-L$  на під-

ставі масиву початкової НВ типу (1) та визначити значення його структурних параметрів  $p$  (тобто  $F(L(p, x_i), f_R(x_i)) \rightarrow opt$ ).

Нехай задана деяка навчальна вибірка (НВ) загального вигляду (1) та деяка система (набір) елементарних ознак (атрибутів) для масиву початкової вибірки  $\phi_1(x), \phi_2(x), \dots, \phi_n(x)$ . Введемо такі множини:

$$G_{r_1, \dots, r_n} = \{x \in G / a_1(x) = r_1, \dots, a_n(x) = r_n\} . \quad (2)$$

Відзначимо, що система множин  $G_{r_1, \dots, r_n}$  представляє повне розбиття множини  $G$ , яке реалізується елементарними ознаками  $\phi_1, \phi_2, \dots, \phi_n$ . Зауважимо, що деякі з множин  $G_{r_1, \dots, r_n}$  можуть бути пустими.

Через  $S_{r_1, \dots, r_n}^j$  позначимо кількість входжень в початкову вибірку тих пар  $(x_i, f_R(x_i))$ ,  $(1 \leq i \leq m)$ , які задовольняють умовам  $x_i \in G_{r_1, \dots, r_n}$ .

Аналогічно через  $S_{r_1, \dots, r_n}^j$ ,  $(j = \overline{0, k-1})$  позначимо кількість входжень в НВ тих пар  $(x_i, f_R(x_i))$ ,  $(i = 1, 2, \dots, m)$ , які задовольняють умовам  $x_i \in G_{r_1, \dots, r_n}$  та  $f_R(x_i) = j$ .

Введемо такі величини:

$$\delta_{r_1, \dots, r_n}^j = \frac{S_{r_1, \dots, r_n}^j}{m}; \quad \psi_{r_1, \dots, r_n}^j = \frac{S_{r_1, \dots, r_n}^j}{S_{r_1, \dots, r_n}}; \quad \rho_{r_1, \dots, r_n} = \max_j \psi_{r_1, \dots, r_n}^j . \quad (3)$$

Зауважимо, що якщо  $x_i \notin G_{r_1, \dots, r_n}$  для всіх  $i = 1, \dots, m$ , тоді  $\delta_{r_1, \dots, r_n}^j = 0$  та  $\psi_{r_1, \dots, r_n}^j = 0$ ,  $(j = \overline{0, 1, \dots, k-1})$ . Величина  $\delta_{r_1, \dots, r_n}^j$  характеризує частоту входжень членів послідовності  $x_1, x_2, \dots, x_m$  в множину  $G_{r_1, \dots, r_n}$ . Величина  $\psi_{r_1, \dots, r_n}^j$  характеризує частоту приналежності об'єкту  $x$  образу (класу)  $H_j$  при умові, що  $x \in G_{r_1, \dots, r_n}$ . Тут варто зауважити, що умова  $x \in G_{r_1, \dots, r_n}$  еквівалентна умові елементарних ознак:

$$\phi_1(x) = r_1, \phi_2(x) = r_2, \dots, \phi_n(x) = r_n .$$

Величина  $\delta_{r_1, \dots, r_n}^j$  характеризує інформаційну ефективність розпізнавання приналежності  $x$  до одного з класів  $H_0, H_1, \dots, H_{k-1}$  звичайно при умов базової умови:

$$x \in G_{r_1, \dots, r_n} .$$

На наступному етапі, нехай  $x \in G_{r_1, \dots, r_n}$ . Тоді виникає питання, до якого класу з  $H_0, H_1, \dots, H_{k-1}$  варто віднести  $x$ . Природно  $x$  варто віднести до того класу  $H_l$  для якого виконується наступне співвідношення:

$$\rho_{r_1, \dots, r_n} = \psi_{r_1, \dots, r_n}^l, \{0 \leq l \leq k-1\} . \quad (4)$$

Відзначимо, що дане співвідношення представляє деяке фінальне правило класифікації. Зрозуміло, що чим більше  $\rho_{r_1, \dots, r_n}$ , тим вища ефективність його загальної ефективності.

Як було вказано вище, в якості єдиної інформації, яка представляє образи  $H_0, H_1, \dots, H_{k-1}$ , мається на увазі початкова НВ. Тому під класом  $H_l$  розуміється сукупність всіх пар  $(x_i, f_R(x_i))$  НВ, які задовольняють базово-му співвідношенню  $f_R(x_i) = l$ .

Середня ефективність розпізнавання образів  $H_0, H_1, \dots, H_{k-1}$ , які задані даними НВ за допомогою елементарних ознак  $\phi_1, \phi_2, \dots, \phi_n$  оцінюється такою величиною:

$$O_S(\phi_1, \phi_2, \dots, \phi_n) = \sum_{r_1, \dots, r_n} \delta_{r_1, \dots, r_n} * \rho_{r_1, \dots, r_n}. \quad (5)$$

Величину  $O_S(\phi_1, \phi_2, \dots, \phi_n)$  можна рахувати оцінкою апроксимації НВ за допомогою набору елементарних ознак  $a_1, a_2, \dots, a_n$ . З виразу (5) випливає таке співвідношення для  $\delta_{r_1, \dots, r_n}$ ,  $\psi_{r_1, \dots, r_n}^j$  та  $\rho_{r_1, \dots, r_n}$ :

$$\delta_{r_1, \dots, r_n} > 0, \sum_{0 \leq r_1, \dots, r_n \leq 1} \delta_{r_1, \dots, r_n} = 1; \psi_{r_1, \dots, r_n}^j \geq 0, \sum_{j=0}^{k-1} \psi_{r_1, \dots, r_n}^j = 1; \quad (6)$$

$$\frac{1}{k} \leq \rho_{r_1, \dots, r_n} \leq 1, (r_1, \dots, r_n \in \{0, 1\}).$$

З формули (6) безпосередньо впливають такі властивості величини:

$$1) \frac{1}{k} \leq O_S(\phi_1, \phi_2, \dots, \phi_n) \leq 1; \quad (7)$$

$$2) O_S(\phi_1, \phi_2, \dots, \phi_n) = 1 \Leftrightarrow \rho_{r_1, \dots, r_n} = 1.$$

Зауважимо, що дані властивості актуальні для всіх  $r_1, \dots, r_n$ , які задовольняють співвідношенню  $\exists i(1 \leq i \leq m, x_i \in G_{r_1, \dots, r_n})$ . З другої частини (7) випливає, що набір елементарних ознак  $\phi_1, \dots, \phi_n$  тоді і тільки тоді реалізує повне розпізнавання образів  $H_0, H_1, \dots, H_{k-1}$  (які задані початковою НВ). Тобто набір елементарних ознак буде тестом при умові  $O_S(\phi_1, \phi_2, \dots, \phi_n) = 1$ , а формула (5) дає змогу знаходити такі набори алгоритмів.

Розглянемо ще одну важливу особливість функціональної оцінки набору елементарних ознак відносно початкової НВ. Нехай маємо деяке число  $b, (1/k \leq b < 1)$ , а  $M_b$  – кількість всіх входжень об'єктів  $x_i$  в послідовність  $x_1, x_2, \dots, x_m$ , для яких виконуються співвідношення:

$$\begin{cases} x_i \in G_{r_1, \dots, r_n}; \\ \rho_{r_1, \dots, r_n} > b. \end{cases} \quad (8)$$

Дане співвідношення фактично представляє ефективність розпізнавання для об'єктів  $x_i$  за допомогою деякого набору елементарних ознак  $\phi_1, \dots, \phi_n$  більше чим  $b$ .

Число  $\gamma_b = m_b / m$  є часткою тих входжень об'єктів  $x_i$  в послідовність  $x_1, x_2, \dots, x_m$  для яких ефективність розпізнавання за допомогою набору елементарних ознак  $\phi_1, \phi_2, \dots, \phi_n$  більше чим  $b$ .

Число  $\gamma_b$  виражається так:

$$\gamma_b = \sum_{\rho_{r_1, \dots, r_n} > b} \delta_{r_1, \dots, r_n}. \quad (9)$$

Зауважимо, що вираз  $\sum_{\rho_{r_1, \dots, r_n} > b} \delta_{r_1, \dots, r_n}$  означає сумування по всіх  $r_1, \dots, r_n$ , які задовольняють співвідношенню  $\rho_{r_1, \dots, r_n} > b$ . Припустимо, що  $O_S(\phi_1, \phi_2, \dots, \phi_n) = c$  та  $1/k \leq b < c$ . Тоді з виразу (5) та співвідношення  $O_S(\phi_1, \phi_2, \dots, \phi_n) = c$  будемо мати:

$$c = \sum_{\rho_{r_1, \dots, r_n} > b} \delta_{r_1, \dots, r_n} \cdot \rho_{r_1, \dots, r_n} + \sum_{\rho_{r_1, \dots, r_n} \leq b} \delta_{r_1, \dots, r_n} \cdot \rho_{r_1, \dots, r_n}. \quad (10)$$

З виразів (6) та (9) отримаємо наступне:

$$\sum_{\rho_{r_1, \dots, r_n} \leq b} \delta_{r_1, \dots, r_n} = 1 - \gamma_b. \quad (11)$$

З виразів (6), (9), (10) та (11) будемо мати:

$$c \leq \gamma_b + (1 - \gamma_b) * b. \quad (12)$$

З виразу (12) безпосередньо випливає:

$$\gamma_b \geq \frac{c - b}{1 - b}. \quad (13)$$

Внесемо у вираз (13)  $c = 1 - \varepsilon^2$  та  $b = 1 - \varepsilon$ , де  $\varepsilon$  – нескінченно мале число. За таких умов будемо мати:

$$\gamma_b \geq 1 - \varepsilon. \quad (14)$$

З виразу (14) випливає, що якщо  $c \rightarrow 1$ , то існує така величина  $b$  (яка залежить від  $c$ ), що:

$$\frac{1}{k} \leq b < c, \lim_{c \rightarrow 1} b = 1.$$

Зауважимо, що величина  $b$  в даному випадку не залежить від початкової НВ (а отже та від її потужності – числа  $m$ ).

Відзначимо, що головна ідея методів поетапної селекції елементарних ознак, полягає в тому щоби максимізувати величину якості ознаки  $W_M(f)$  [34]. Останнє означає, що а алгоритмах логічного дерева має бути знайдена для навчальної вибірки (1) така узагальнена ознака  $f$ , для якої величина  $W_M(f)$  є за можливості найбільшою.

Відзначимо, що під важливістю елементарної ознаки (інформативністю) будемо розуміти величину, яка може бути розрахована, як варіант, наступними функціоналами:

$$W(\phi) = \frac{1}{M} \sum_{i=1}^M b_i \cdot \rho_i, \text{ де } \rho_i = \max_{1 \leq m \leq k} \frac{q_i^m}{b_i}; \quad (15)$$

$$W(\phi) = \frac{1}{M} \sum_{j \in G_i} \max_{0 \leq k \leq l} b_j^k. \quad (16)$$

Зрозуміло, що аналогічно можна оцінити важливість інших ознак. Величину  $q_i^m / b_i$  можна інтерпретувати як імовірність того, що функція  $f_R(x)$  прийме значення  $O_m, (1 \leq m \leq k)$ , при умові, що значення ознаки  $\phi$  дорівнює  $i, (1 \leq i \leq k)$ . Величина  $\rho_i$  представляє максимальну з цих ймовірностей. Можна сказати, що величина  $\rho_i$  представляє ту інформацію, яку можна отримати про значення функції  $f_R(x)$ , знаючи, що на наборі  $z$  значення ознаки  $\phi$  дорівнює  $i$ . Тут величина  $W(f)$ , яка визначається даною формулою, характеризує ту інформацію, яку можна отримати про функцію  $f_R(x)$ , якщо відомо значення ознаки  $\phi$  на наборі  $z$ .

Зрозуміло, що елементарна ознака, для якої ця інформація є найбільшою, вважається найбільш важливою ознакою за відношенням до  $f_R(x)$ . Відповідно  $G_i$  – множина значень  $\phi_i$ -ої ознаки,  $b_j^k$  – кількість значень  $j$ -ої ознаки в класі  $H_k, (k = 1, \dots, l)$ ,  $l$  – кількість класів  $H$ .

Відзначимо, що по аналогії з деревами рішень – функціонали (15) в даній схемі побудови ЛДК будуть виступати базовим критерієм розгалуження самої структури логічного дерева.

Отже, в даному методі побудови ЛДК спочатку вибирається деяка елементарна ознака  $\phi_i^1$ . Від цієї елементарної ознаки вимагається, щоби величина  $W_M(\phi_i^1)$ , яка характеризує фактично її інформативність відносно початкової вибірки (1), була за можливості найбільшою, що практично легко досягти.

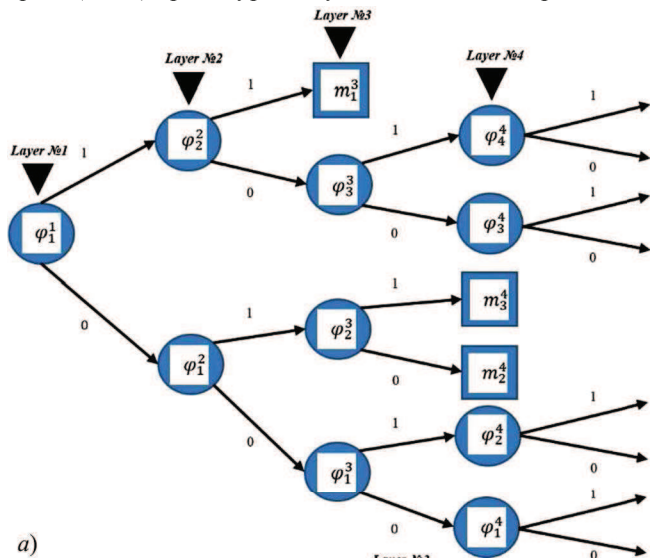
Зауважимо же раз, що  $W_M(\phi_i^1)$  розраховується відповідно до методики (15). Наступні кроки методу побудови логічного дерева зручно інтерпретувати за допомогою наступної схеми – (рис. 1, а).



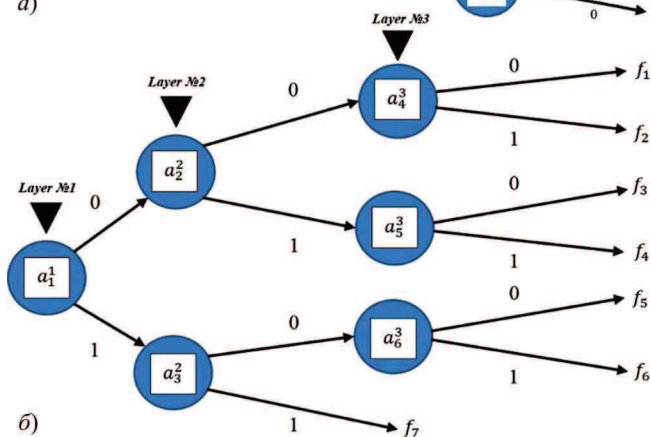
Відзначимо, що в кожній вершині ЛДК – (рис. 1,а) знаходиться або деяка ознака  $\phi_i^j$  (мітка, атрибут), або число  $m_i^j$ , яке належить множині  $\{0,1,\dots,k-1\}$ . Так вершина, в якій знаходиться  $m_i^j$ , називається кінцевою вершиною дерева ЛДК (таку вершину ще називають листом логічного дерева).

Від кожної вершини, в якій знаходиться ознака  $\phi_i^j$ , відходять дві направляючі (стрілки) які позначені 0 та 1. Направляючий, яка позначена 0, відповідає значення  $\phi_i^j = 0$ , а позначеній 1 – значення  $\phi_i^j = 1$ . Логічне дерево розбите умовно за ярусами (рівнями), причому в  $j$ -му ярусі ЛДК стоять відповідні елементарні ознаки  $\phi_1^j, \phi_2^j, \dots$ .

Відзначимо, що всі елементарні ознаки, які стоять в усіх ярусах логічного дерева, починаючи з першого та закінчуючи  $n$ -им, фактично представляють собою ті ознаки, які отримані після проведення  $n$  кроків (етапів) процесу побудови дерева класифікації (ЛДК). Причому елементарні ознаки, які стоять на  $n$ -му ярусі, представляють собою ті ознаки, які отримані відповідно на  $n$ -му кроці (етапі) процедури побудови логічного дерева.



а)



б)

Рис. 1. Схема структури логічного дерева класифікації (а), його завершена структура (б)

Відзначимо, що дерево класифікації (рис. 1,а) фактично реалізує (представляє) деяку узагальнену ознаку  $f_i(x)$ , яка визначена на множині  $G$  та приймає значення з множини  $\{0,1,\dots,k-1\}$ . Після побудови узагальненої ознаки  $f_i(x)$  починається етап перевірного тесту.

Так, в режимі перевірного тесту підраховується загальна кількість  $S$  всіх тих пар  $(x_i, f_R(x_i))$  з вибірки типу (1) для яких виконується співвідношення  $f_R(x_i) = f_i(x)$ .

На наступному етапі необхідно перевірити умову  $s/M \geq \delta$ . Відзначимо, що тут  $\delta$  – параметр, який характеризує оцінку ефективності навчання, відносно поточної задачі (НВ). Якщо ця умова виконана – то на цьому процес побудови дерева класифікації закінчується, а узагальнена ознака  $f_i(x)$ , яка видається логічним деревом структури – (рис. 1,а) є такою що забезпечує апроксимацію вибірки вигляду (1).

Звернемо увагу, що в такій схемі методу побудови логічного дерева класифікації є принциповий недолік, який пов'язаний з тим що зі зростанням кількості вершин в логічному дереві (фактично ярусів ЛДК) кількість елементарних ознак  $\phi_i^j$  (тут  $i$  – номер елементарної ознаки в наборі,  $j$  – номер ярусу розташування атрибуту) в дереві значно збільшується. Звичайно таке ускладнення результуючого дерева класифікації негативно впливає на апаратні можливості системи розпізнавання (пам'ять, процесорний час).

Для того, щоби подолати ці негативні моменти можна запропонувати наступну модифікацію методу апроксимації даних НВ набором елементарних ознак. Зафіксуємо деяке додатне число  $z$ . Нехай маємо логічне дерево наступної структури – (рис. 1,б), яке відображає деякий предикат (узагальнену ознаку)  $f_i(x)$ .

Так на етапі тесту обчислювали деяке число  $S$ , яке фігурує у співвідношенні  $s/M \geq \delta$ . Тепер, окрім числа  $S$ , для кожного незавершеного шляху  $r_1, r_2, r_3$  логічного дерева (рис. 1,б) розраховуємо ще число  $S_{r_1, r_2, r_3}$ , де  $S_{r_1, r_2, r_3}$  – кількість всіх пар  $(x_i, f_R(x_i))$  з НВ, які фактично належать шляху  $r_1, r_2, r_3$  і для яких виконується співвідношення  $f_R(x_i) \neq l(r_1, r_2, r_3)$ . Отже,  $S_{r_1, r_2, r_3}$  – це число всіх тих помилок, які здійснюється деяким предикатом  $f_i(x)$  (узагальненою ознакою), яке видається ЛДК (рис. 1,б) на шляху  $r_1, r_2, r_3$ .

На наступному кроці вибираємо число  $S$  таких шляхів  $(r_1, r_2, r_3)_1, \dots, (r_1, r_2, r_3)_z$ , для яких число  $S_{r_1, r_2, r_3}$  буде найбільшим. Наприклад, нехай  $z = 3$ , та має місце співвідношення  $S_{000} \geq S_{100} \geq S_{101} \geq S_{001} \geq S_{010} \geq S_{011}$ . Тоді вибираються тільки шляхи 000,100,101. Наступна добудова відбір вершин (елементарних ознак)  $\phi_{r_1, r_2, r_3}$  здійснюється тільки для цих шляхів.

Дану схему побудови дерева класифікації будемо називати обмеженим методом побудови структури ЛДК тому що за такою схемою, в процесі побудови дерева класифікації продовжуються тільки ті шляхи за якими відбувається найбільша кількість помилок. Тут варто зауважити, що при застосуванні тільки що вказаного процесу в кінці шляхів  $r_1, r_2, r_3$ , які не входять у відібрані  $z$  шляхів, значення  $l(r_1, r_2, r_3)$  зберігаються.

Зауважимо, що процес обмеженого методу побудови ЛДК можна застосовувати в тому випадку, коли початкова НВ типу (1) не є фіксованою, тобто коли на кожному кроці процесу подається своя вибірка.

**Аналіз останніх досліджень та публікацій.** Дане дослідження продовжує цикл робіт, які присвячені про-

блематиці деревоподібних схем розпізнавання, методам та алгоритмам класифікації дискретних об'єктів [15], [16], [17], [20]. В них піднімаються питання побудови, аналізу та оптимізації логічних дерев класифікації [19], [29], [30], [31]. Так з виразу [2] 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), які групуються по ярусам і які отримані на певному кроці (етапі) побудови дерева розпізнавання [11], [18], [21]. Важливою задачею, яка виникає з [4], [7], [8], [10] задача синтезу дерев розпізнавання, які 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 [1]. Так, в статі [27] розглядається випадок побудови дерев рішень для набору малоінформативних ознак.

Здатність ЛДК 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) [28], [32]. Така концепція логічних дерев активно використовують в інтелектуальному аналізі даних, де кінцева мета полягає в синтезі моделі, яка прогнозує значення цільової змінної на підставі набору початкових даних на вході системи [9]. Оскільки головну ідею методів та алгоритмів розгалуженого вибору ознак можна визначити – як оптимальну апроксимацію деякої початкової НВ набором елементарних ознак (атрибутів об'єкту), то на перший план виходить їх центральна проблема – питання вибору ефективного критерію розгалуження (відбору вершин, атрибутів, ознак дискретних об'єктів) [24], [25], [26]. Саме цим принциповим задачам присвячені роботи [21] де піднімаються питання якісної оцінки окремих дискретних ознак, їх наборів та фіксованих сполучень, що дає змогу запровадити ефективний механізм реалізації розгалуження.

Варто відзначити, що абсолютна більшість з відомих алгоритмів побудови дерев рішень відносяться до класу "жадібних алгоритмів". Тобто, якщо на якомусь етапі була відібрана деяка вершина (атрибут, вузол) та по ній проведене розбиття на підмножини початкової вибірки – алгоритм не може на наступному кроці повернутися назад для вибору іншої вершини (вузла) з більш якісними показниками розбиття. Це фактично означає, що на етапі побудови логічного дерева неможливо визначити чи дасть відібрана вершина розгалуження в кінцевому етапі деяке оптимальне розбиття [13], [14]. Так, при синтезі дерев рішень центральними питаннями залишаються питання вибору критерію атрибуту (вершини ЛДК), за якою відбудеться розбиття початкової НВ, критерій зупинки навчання (побудови структури ЛДК) та критерій відкидання гілок логічного дерева (піддерев ЛДК) [3].

Структури логічних дерев класифікації, які отримані на онові методів розгалуженого вибору ознак, характеризуються компактністю з одного боку та нерівномірністю заповнення (розрядженістю) ярусів з іншого боку порівняно з регулярними деревами (алгоритмом з одно-

рзовою оцінкою важливості ознак) [29]. Відзначимо, що важливими питаннями залишаються питання збіжності процесу побудови ЛДК за методами селекції елементарних ознак та питання вибору критерію зупинки процесу синтезу логічного дерева (наприклад, обмеження за глибиною або складністю логічного дерева, обмеження за точністю або кількістю помилок структури, що будується) [5]. Вирішення даного питання в розрізі обмеженого (модифікованого) методу побудови ЛДК буде представлено в даному дослідженні далі.

Отже, не дивлячись на певну проблематику, яка виникає при побудові та використанні дерев рішень (концепції ЛДК в загальному сенсі) варто зафіксувати такі їх переваги:

- 1) Деревоподібні моделі відрізняють принципово швидкий етап навчання системи розпізнавання.
- 2) Можливість синтезу множини правил рішень в межах області де навіть кваліфікованому експерту важко сформулювати набір рекомендацій.
- 3) Синтез правил (класифікаційних правил, правил рішень) на природній мові.
- 4) Отримана результуюча деревоподібна модель класифікації – інтуїтивно зрозумілою.
- 5) Побудований прогноз, який отримується на останньому етапі роботи моделі, характеризується високою точністю, навіть порівняно з статистичними та нейромережевими моделями.
- 6) Можливість побудови непараметричних моделей.

Отже, зважаючи на все вказане вище, можна зробити висновок, що моделі дерев рішень разом з нейромережним підходом [1], [12] є важливим та актуальним інструментом широкого аналізу та представлення структури даних.

## Результати дослідження та їх обговорення

Відзначимо, що запропонована схема побудови ЛДК дає змогу регулювати складність моделі дерева класифікації, що будується, або будувати модель з наперед заданою точністю. Причому, задача відбору моделі дерева класифікації серед набору побудованих ЛДК для конкретної задачі визначається набором параметрів, які мають визначальну важливість відносно поточної прикладної задачі (набору даних НВ).

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

Принципово на даному етапі дослідження розглянути критерії якості отриманих інформаційних моделей, які залежать від похибки моделі, потужності початкового масиву даних НВ (кількість навчальних пар та розмірність ознакового простору задачі), кількості параметрів моделі і так далі.

Зрозуміло, що критично важливими параметрами побудованої моделі ЛДК, які необхідно мінімізувати є помилки моделі відповідно на масивах даних НВ, ТВ та для кожного з класів, які задані початковою умовою поточної прикладної задачі.

Зауважимо, що принциповим моментом залишається питання зменшення складності структури ЛДК (мається на увазі кількість ознак в структурі ЛДК, загальна кількість вершин моделі ЛДК та загальна кількість переходів в структурі ЛДК), параметри загальних витрат

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

$$Q_{Main} = \frac{Fr_{All}}{V_{All} \cdot \sum_i p_i} \cdot \text{Exp} \left( -\frac{Er_{All}}{M_{All}} \right). \quad (16)$$

Відзначимо, що тут параметр  $Er_{All}$  – загальна кількість помилок моделі ЛДК на масивах даних початкових тестової та навчальної вибірки, відповідно  $M_{All}$  – загальна потужність цих двох масивів даних. Параметр  $Fr_{All}$  – характеризує кількість вершин отриманої моделі ЛДК з результируючими значеннями  $f_r$  (ФР, тобто листів дерева класифікації), а параметр  $V_{All}$  – представляє загальну кількість всіх типів вершин в структурі моделі ЛДК. Набір параметрів  $p_i$  представляє найбільш важливі характеристики дерева класифікації, що оцінюється (кількість елементарних ознак що використовують в моделі дерева класифікації, кількість переходів між вершинами, ярусами дерева класифікації, тощо).

Відзначимо, що даний інтегральний показник якості моделі ЛДК буде приймати значення в межах нуля та одиниці. Чим менший він буде тим гірша буде якість побудованого дерева класифікації, а чим більший буде показник тим краще буде отримана модель.

Запропонований обмежений метод побудови моделей ЛДК (модифікований метод селекції елементарних ознак) пройшов апробацію задачею класифікації масиву геологічних даних (для нього була сконструйована автономної системи розпізнавання на підставі деревоподібної моделі класифікації) – задача про розділення нафтоносних пластів. Початкові параметри даної прикладної задачі представлені в наступній таблиці.

Табл. 1. Початкові параметри задачі

Розмірність ознакового простору $N$	Потужність масиву даних початкової НВ – $M$	Потужність масиву даних ТВ – $S$	Загальна кількість класів за розбиттям даних НВ –	Відношення об'єктів різних класів НВ – $(H_1 / H_2)$
22	1250	240	2	756/494

Так, у НВ представлена інформація про розбиття на два класи. На етапі екзамени побудована система класифікації має забезпечити ефективне розпізнавання об'єктів невідомої класифікації відносно цих двох класів. Зауважимо, що на початковому етапі навчальна та тестова вибірка була автоматично перевірена на коректність (пошук та видалення однакових об'єктів різної належності – помили першого та другого роду).

**Обговорення результатів дослідження.** В масиві навчальної інформації переважали навчальні пари класу  $H_1$  (нафтоносні пласти) в пропорції  $\approx (1,5 / 1)$ , а сама НВ складалася з 1250 об'єктів, причому ефективність сконструйованої системи розпізнавання оцінювалася на тестовій вибірці об'єму 240 об'єктів, причому ТВ представляла собою відокремлену частину НВ (складалася з дискретних об'єктів відомої класифікації). Дані масивів навчальних та тестових вибірок отримані на підставі геологічної розвідки на території Закарпатської області в період з 2001 року по 2011 рік.

Так, фрагмент основних результатів, наведених вище експериментів (порівняльних тестів методів побудови моделей ЛДК на масиві даних даної прикладної задачі), представлений в (табл. 2). Причому побудовані моделі ЛДК забезпечили необхідний рівень точності заданий умовою задачі, швидкодію та витрати робочої пам'яті системи, але показували різну структурну складність побудовах дерев класифікації та набору узагальнених ознак (у випадку моделі алгоритмічного дерева класифікації [20]).

Зауважимо, що запропонована в дослідженні оцінка якості моделі дерева класифікації відображає базові параметри (характеристики) дерев класифікації та може бути застосована в якості критерію оптимальності в процедурі оцінки довільної деревоподібної схеми розпізнавання, наприклад у випадку методів побудови та відбору випадкових ЛДК з роботи [21].

Табл. 2. Порівняльна таблиця моделей / методів дерев класифікації

№ п/п	Метод синтезу структури логічного дерева	Інтегральний показник якості моделі $Q_{Main}$	Кількість помилок моделі на НВ та ТВ $Er_{All}$
1	Метод повного ЛДК на підставі селекції елементарних ознак	0,004789	2
2	Модель ЛДК з одноразовою оцінкою важливості ознак	0,002263	3
3	Обмежений метод побудови ЛДК	0,003181	2
4	Метод алгоритмічного дерева (типу I)	0,005234	0
5	Метод алгоритмічного дерева (типу II)	0,002941	0

Так, запропонований в роботі обмежений метод селекції елементарних ознак (модифікований метод побудови ЛДК) порівнювався з методом повного ЛДК, алгоритмом з одноразовою (початковою) оцінкою важливості дискретних ознак, методом алгоритмічного дерева класифікації та показав загалом прийнятний результат.

На завершення відзначимо, що головна ідея методів дерев класифікації на підставі автономних алгоритмів в своїй структурі, полягає в поетапній апроксимації відібраним набором алгоритмів масиву даних початкової НВ (що відображається в конструкції самого дерева). Так, отримана структура АДК з одного боку характеризується високою універсальністю відносно прикладних задач та відносно компактною структурою самої моделі, але з іншого боку вимагає істотних апаратних витрат (пам'ять та процесорний час) для зберігання узагальнених ознак (параметрів правил класифікації) та початкової оцінки якості зафіксованих алгоритмів класифікації за даними НВ. Тому порівняно з концепцією АДК методи ЛДК (обмежений метод селекції елементарних ознак) має високу швидкодію схем класифікації, відносно незначні апаратні витрати для зберігання та роботи самої структури дерева (структур даних) та високу (регульовану складність) якість класифікації дискретних об'єктів.



## Висновки

Отже, загалом вирішене завдання автоматизації побудови ЛДК на підставі обмеженого методу апроксимації НВ набором елементарних ознак та запропонована схема побудови дерева класифікації з наперед заданою точністю. В роботі розроблений обмежений метод побудови моделі ЛДК, який дає змогу будувати дерева класифікації заданої точності регулюючи складність схеми яка генерується.

Такий підхід дає змогу добитися оптимальної структури моделі ЛДК, що будується та забезпечити необхідну і достатню точність отриманого дерева класифікації. Критерії розгалуження в структурі ЛДК, які представлені в роботі можна ефективно застосувати не тільки для оцінки інформативності окремих елементарних ознак, але й для розрахунку важливості наборів ознак та їх сполучень, що дає змогу піднімати питання особливості побудови молей ЛДК для мало інформативних ознак [27].

Також запропонований загальний інтегральний показник якості моделі ЛДК, який дає змогу ефективно подати загальні характеристики дерев класифікації, причому можливе його використання для відбору найбільш оптимального ЛДК у випадку методів побудови випадкових ЛДК [21].

## References

- [1] Bodyanskiy, Y., Vynokurova, O., Setlak, G. & Pliss, I. (2015). *Hybrid neuro-neo-fuzzy system and its adaptive learning algorithm*, Xth Scien. and Tech. Conf. "Computer Sciences and Information Technologies" (CSIT), Lviv, 111-114. <https://doi.org/10.1109/STC-CSIT.2015.7325445>
- [2] Breiman, L. L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Boca Raton, Chapman and Hall/CRC, 368 p.
- [3] De Mántaras, R. L. (1991). A distance-based attribute selection measure for decision tree induction. *Machine learning*, 6(1), 81–92. <https://doi.org/10.1007/BF00153761>
- [4] Deng, H., Runger, G., & Tuv, E. (2011). Bias of importance measures for multi-valued attributes and solutions, *Proceedings of the 21st International Conference on Artificial Neural Networks (ICANN)*, 293–300. [https://doi.org/10.1007/978-3-642-21738-8\\_38](https://doi.org/10.1007/978-3-642-21738-8_38)
- [5] Deng, H., Runger, G., & Tuv, E. (2011). *Bias of importance measures for multi-valued attributes and solutions*, 21st International Conference on Artificial Neural Networks (ICANN), Espoo, 14–17 June 2011: proceedings. Berlin, Springer-Verlag, 2, 293–300. [https://doi.org/10.1007/978-3-642-21738-8\\_38](https://doi.org/10.1007/978-3-642-21738-8_38)
- [6] Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning*. Berlin, Springer, 768.
- [7] Kamiński, B., Jakubczyk, M., & Szufel, P. (2017). [A framework for sensitivity analysis of decision trees](https://doi.org/10.1007/s10100-017-0479-6). *Central European Journal of Operations Research*, 26(1), 135–159. <https://doi.org/10.1007/s10100-017-0479-6>
- [8] Karimi, K. L., & Hamilton, H. J. (2011). Generation and Interpretation of Temporal Decision Rules. *International Journal of Computer Information Systems and Industrial Management Applications*, 3, 314–323.
- [9] Koskimaki, H., Juutilainen, I., Laurinen, P., & Roning, J. (2008). *Two-level clustering approach to training data instance selection: a case study for the steel industry*, *Neural Networks: International Joint Conference (IJCNN-2008)*, Hong Kong, 1–8 June 2008: proceedings. Los Alamitos, IEEE, 3044–3049. <https://doi.org/10.1109/IJCNN.2008.4634228>
- [10] Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatica*, 31, 249–268.
- [11] Laver, V. O., & Povkhan, I. F. (2019). The algorithms for constructing a logical tree of classification in pattern recognition problems. *Scientific notes of the Tauride national University. Series: technical Sciences*, 30(69), 4, 100–106. <https://doi.org/10.32838/2663-5941/2019.4-1/18>
- [12] Lupei, M., Mitsa, A., Repariuk, V., & Sharkan, V. (2020). Identification of authorship of Ukrainian-language texts of journalistic style using neural networks. *Eastern-European Journal of Enterprise Technologies*, 2(103), 30–36. <https://doi.org/10.15587/1729-4061.2020.195041>
- [13] Miyakawa, M. (1989). Criteria for selecting a variable in the construction of efficient decision trees. *IEEE Transactions on Computers*, 38(1), 130–141. <https://doi.org/10.1109/12.8736>
- [14] Painsky, A., & Rosset, S. (2017). Cross-validated variable selection in tree-based methods improves predictive performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(11), 2142–2153. <https://doi.org/10.1109/TPAMI.2016.2636831>
- [15] Povhan, I. (2016). *Designing of recognition system of discrete objects*, IEEE First International Conference on Data Stream Mining & Processing (DSMP), Ukraine. Lviv, 226–231.
- [16] Povhan, I. (2019). General scheme for constructing the most complex logical tree of classification in pattern recognition discrete objects. *Collection of scientific papers Electronics and information technologies*, Lviv, 11, 112–117. <https://doi.org/10.30970/eli.11.7>
- [17] Povhan, I. F. (2019). The problem of general estimation of the complexity of the maximum constructed logical classification tree. *Bulletin of the national technical University Kharkiv Polytechnic Institute*, 13, 104–117. <https://doi.org/10.20998/2411-0558.2019.13.10>
- [18] Povkhan, I. (2020). Classification models of flood-related events based on algorithmic trees. *Eastern-European Journal of Enterprise Technologies*, 6-4(108), 58–68. <https://doi.org/10.15587/1729-4061.2020.219525>
- [19] Povkhan, I. F. (2018). The problem of functional evaluation of a training sample in discrete object recognition problems. *Scientific notes of the Tauride national University. Series: technical Sciences*, 29(68), 6, 217–222.
- [20] Povkhan, I. F. (2019). *Features of synthesis of generalized features in the construction of recognition systems using the logical tree method*, Materials of the international scientific and practical conference "Information technologies and computer modeling ITKM-2019". Ivano-Frankivsk, 169–174.
- [21] Povkhan, I. F. (2019). Features random logic of the classification trees in the pattern recognition problems. *Scientific notes of the Tauride national University. Series: technical Sciences*, 30(69), 5, 152–161. <https://doi.org/10.32838/2663-5941/2019.5-1/22>
- [22] Quinlan, J. R. (1986). Induction of Decision Trees, *Machine Learning*, 1, 81–106. <https://doi.org/10.1007/BF00116251>
- [23] Srikant, R., Agrawal, R. (1997). Mining generalized association rules. *Future Generation Computer Systems*, 13(2), 61–180. [https://doi.org/10.1016/S0167-739X\(97\)00019-8](https://doi.org/10.1016/S0167-739X(97)00019-8)
- [24] Subbotin, S. (2013). The neuro-fuzzy network synthesis and simplification on precedents in problems of diagnosis and pattern recognition. *Optical Memory and Neural Networks (Information Optics)*, 22(2), 97–103. <https://doi.org/10.3103/S1060992X13020082>
- [25] Subbotin, S. A. (2013). Methods of sampling based on exhaustive and evolutionary search. *Automatic Control and Computer Sciences*, 47(3), 113–121. <https://doi.org/10.3103/S0146411613030073>
- [26] Subbotin, S. A. (2014). Methods and characteristics of localitypreserving transformations in the problems of computational intelligence. *Radio Electronics, Computer*

- Science, Control, 1*, 120–128. <https://doi.org/10.15588/1607-3274-2014-1-17>
- [27] Subbotin, S. A. (2019). Construction of decision trees for the case of low-information features. *Radio Electronics, Computer Science, Control, 1*, 121–130. <https://doi.org/10.15588/1607-3274-2019-1-12>
- [28] Subbotin, S., & Oliinyk, A. (2017). The dimensionality reduction methods based on computational intelligence in problems of object classification and diagnosis, Recent Advances in Systems, Control and Information Technology, [eds.: R. Szewczyk, M. Kaliczyńska]. Cham, Springer, 11–19. [https://doi.org/10.1007/978-3-319-48923-0\\_2](https://doi.org/10.1007/978-3-319-48923-0_2)
- [29] Vasilenko, Y. A., Vashuk, F. G., & Povkhan, I. F. (2011). The problem of estimating the complexity of logical trees recognition and a general method for optimizing them. *European Journal of Enterprise Technologies, 6/4(54)*, 24–28.
- [30] Vasilenko, Y. A., Vashuk, F. G., & Povkhan, I. F. (2012). General estimation of minimization of tree logical structures. *European Journal of Enterprise Technologies, 1/4(55)*, 29–33.
- [31] Vasilenko, Y. A., Vashuk, F. G., Povkhan, I. F., Kovach, M. Y., & Nikarovich, O. D. (2004). Minimizing logical tree structures in image recognition tasks. *European Journal of Enterprise Technologies, 3(9)*, 12–16.
- [32] Vasilenko, Y. A., Vasilenko, E. Y., & Povkhan, I. F. (2002). Defining the concept of a feature in pattern recognition theory. *Artificial Intelligence, 4*, 512–517.
- [33] Vasilenko, Y. A., Vasilenko, E. Y., & Povkhan, I. F. (2003). Branched feature selection method in mathematical modeling of multi-level image recognition systems. *Artificial Intelligence, 7*, 246–249.
- [34] Vasilenko, Y. A., Vasilenko, E. Y., & Povkhan, I. F. (2004). Conceptual basis of image recognition systems based on the branched feature selection method. *European Journal of Enterprise Technologies, 7(1)*, 13–15.

**I. F. Povkhan**

*Uzhhorod National University, Uzhhorod, Ukraine*

## THE METHOD OF BOUNDED CONSTRUCTIONS OF LOGICAL CLASSIFICATION TREES IN THE PROBLEM OF DISCRETE OBJECTS CLASSIFICATION

The problem of constructing a model of logical classification trees based on a limited method of selecting elementary features for geological data arrays is considered. A method for approximating an array of real data with a set of elementary features with a fixed criterion for stopping the branching procedure at the stage of constructing a classification tree is proposed. This approach allows to ensure the necessary accuracy of the model, reduce its structural complexity, and achieve the necessary performance indicators. A limited method for constructing classification trees has been developed, which is aimed at completing only those paths (tiers) of the classification tree structure where there are the greatest number of errors (of all types) of classification. This approach to synthesizing the recognition model makes it possible to effectively regulate the complexity (accuracy) of the classification tree model that is being built, and it is advisable to use it in situations with restrictions on the hardware resources of the information system, restrictions on the accuracy and structural complexity of the model, restrictions on the structure, sequence and depth of recognition of the training sample data array. The limited scheme of synthesis of classification trees allows to build models almost 20 % faster. The constructed logical classification tree will accurately classify (recognize) the entire training sample that the model is based on, will have a minimal structure (structural complexity), and will consist of components – sets of elementary features as design vertices, tree attributes. Based on the proposed modification of the elementary feature selection method, software has been developed that allows working with a set of different types of applied problems. An approach to synthesizing new recognition models based on a limited logic tree scheme and selecting pre-pruning parameters is proposed. In other words, an effective scheme for recognizing discrete objects has been developed based on step-by-step evaluation and selection of sets of attributes (generalized features) based on selected paths in the classification tree structure at each stage of scheme synthesis.

**Keywords:** logical classification tree; image recognition; classification; attribute; branching criterion.

### Інформація про автора:

**Повхан Ігор Федорович**, канд. техн. наук, доцент, кафедра програмного забезпечення. **Email:** povkhan.igor@uzhnu.edu.ua; <https://orcid.org/0000-0002-1681-3466>

**Цитування за ДСТУ:** Повхан І. Ф. Метод обмежених структур логічних дерев у задачі класифікації дискретних об'єктів. Український журнал інформаційних технологій. 2021, т. 3, № 1. С. 22–29.

**Citation APA:** Povkhan, I. F. (2021). The method of bounded constructions of logical classification trees in the problem of discrete objects classification. *Ukrainian Journal of Information Technology, 3(1)*, 22–29. <https://doi.org/10.23939/ujit2021.03.022>