Вісник Національного технічного університету "ХПІ", 2019, № 28 (1353)
ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

УДК 004.8: 004.89: 519.7 DOI: 10.20998/2411-0558.2019.28.05

І. Ф. ПОВХАН, канд. техн. наук, доц., ДВНЗ "Ужгородський національний університет", Ужгород

ПИТАННЯ ПОБУДОВИ ДЕРЕВОПОДІБНИХ МОДЕЛЕЙ РОЗПІЗНАВАННЯ ОБРАЗІВ

Робота піднімає важливі питання теорії розпізнавання образів (дискретних об‘єктів), які пов‘язані з загальною проблематикою побудови деревоподібних схем розпізнавання, класифікації та плавно підводить до концепції алгоритмічного дерева класифікації, що безумовно є найвищим рівнем абстракції в теорії дерев розпізнавання. Простий, ефективний, економний метод побудови логічного дерева класифікації навчальної вибірки дозволяє забезпечити необхідну швидкодію, рівень складності схеми розпізнавання, що гарантує проведення простого та повного розпізнавання дискретних об‘єктів. На сьогоднішній день існують різноманітні методи побудови як логічних дерев з одноразовим використанням ознак в структурі логічного дерева (алгоритми випадкових дерев, метод розгалуженого вбору ознак з початковою оцінкою інформативності), так і дерев з повторами різних ознак на ярусах логічного дерева (алгоритм побудови дерева з покроковою оцінкою важливості ознак, тощо). В роботі фіксуються суттєві переваги логічних дерев класифікації – програмна простота побудови дерева класифікації, зменшення часу загальної генерації логічного дерева та інше. Іл.: 2. Бібліогр.: 15 назв.

Ключові слова: теорія розпізнавання образів; логічне дерево; алгоритмічні дерева класифікації; навчальна вибірка.

Постановка проблеми та аналіз літературних джерел. Як відомо, вибір системи опорних множин представляється першим і основним етапом в загальній схемі алгоритмів обчислення оцінок, детально опрацьованих ще Ю. Журавльовим, який в свій час запропонував більш загальну формалізацію з різними способами вибору інформативних підсистем ознак та формулами обчислення оцінок. Зрозуміло, що вибір системи опорних множин також визначає множину задач, для розв‘язку яких можна побудувати коректний алгоритм класифікації [1].

Відмітимо, що можливість ефективної реалізації алгоритмів типу обчислення оцінок, істотно залежить від вибору системи опорних множин. Алгоритми обчислення оцінок мають загальну схему, яка складається з послідовності шести етапів, для кожного з яких існують різні варіанти безпосередньої реалізації. В даній роботі велика увага приділена саме першому етапу в схемі даних алгоритмів за рахунок введення нового поняття – поняття *Т*-опорної множини, показаний зв‘язок *Т*-опорних множин з логічних дерев класифікації (ЛДК), та за допомого дерев моделей робиться нульове наближення до концепції алгоритмічного дерева класифікації (АДК).

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

Слід також зауважити, що питання опорних множин безпосередньо пов‘язане з питанням ефективного пошуку тупикових тестів, Саме питання пошуку множини всіх тупикових тестів є обчислювально складна комбінаторна задача, і навіть при сучасному програмно-апаратному забезпеченні (навіть з використанням механізму розпаралелювання) не може бути вирішена навіть для відносно невеликих навчальних вибірках (НВ) (декілька тисяч об‘єктів та ознак). Тому при розв‘язку практичних задач обчислюється та використовується лише певна частка тупикових тестів.

В даному дослідженні вводиться модифікація поняття опорної множини. Істотну роль при цьому грають логічні дерева класифікації, які розглядалися в роботах раніше [2 – 7, 12 – 15]. Проте центр уваги зміщений на вибір систем *T*-опорних множин (головна ідея яких буде представлена нище), що дозволяє значно зменшити час класифікації довільного, допустимого об’єкта *S* за рахунок попередньої обробки начальної інформації (організація асоціативного пошуку). Представлення алгоритмів розпізнавання у вигляді ЛДК дозволяє економити апаратну пам‘ять комп‘ютера при їх практичній реалізації.

Аналізуючи проблематику деревоподібних моделей класифікації та розпізнавання можна побачити певний брак досліджень в цьому напрямку, коли головна увага зміщена в бік концепції нейромережевого розпізнавання. Дане же дослідження продовжує цикл робіт, які присвячені методам та підходам деревоподібних схем розпізнавання (систем класифікації) дискретних об‘єктів [2 – 7]. В них піднімаються питання побудови, використання, та оптимізації логічних дерев. Так з [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), які групуються по ярусам і які отримані на певному кроці (етапі) побудови дерева розпізнавання [4]. Важливою задачею, яка виникає з [5] задача синтезу дерев розпізнавання, які будуть представлятися фактично деревом (графом) алгоритмів.

На відміну від існуючих 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 [2]. Причому слід пам‘ятати, числова величина вказаної важливості характеризує собою помилку розподілу об‘єктів на класи.

Розглянемо модель добре відомих алгоритмів розпізнавання та класифікації типу обчислення оцінок.

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

Нехай стоїть задача класифікації за l класами об‘єктів деякої множини $M \in M_1 \cdot ... \cdot M_n$, де $M_i -$ область ознак (метричний простір з метрикою p_i, $(i = 1,...,n)$).

За допомогою навчаючої інформації $I(l)$ об‘єкта S, будуються оцінки $\Gamma_j(S)$, що належать j-му класу, та за допомогою порогових функцій проводиться класифікація об‘єкта. Кожний конкретний алгоритм класифікації A в моделі визначається вибором системи опорних множин Ω_A, функцією близькості $B(w,S,S^`)$, завданням ваги об’єктів за навчаючою інформацією та правилом класифікації.

Оцінки $\Gamma_j(S)$, як правило, вираховуються за наступною формулою:

$$\Gamma_j(S) = \frac{1}{N*|W|} \sum_{S^` \in W_j} \gamma(S^`) \sum_{w \in \Omega \in \Omega_A} P(w) B(w,S,S^`). \tag{1}$$

Зауважимо, що тут N – множник, що нормує; $|...|$ – потужність (вага) деякої множини; W_j – множина об’єктів навчаючої інформації, що належать j-му класу $(j = 1,...,l)$; $\gamma(S^`)$ – деяке число, що відповідає кожному об’єкту з навчаючої інформації $I(l)$ (характеризує ступінь важливості об‘єкта $S^`$, його інформативність); $P(w) = P_{i_1} + ... + P_{i_k}$ – вага опорної множини $w \leftrightarrow \Omega = \{i_1,...,i_k\}$; $S = (a_1,...,a_n)$, $S' = (b_1,...,b_n)$ – об’єкти з M; $B(w,S,S') = 1$ якщо в деякій системі нерівностей $p_{i_1}(a_{i_1}, b_{i_1}) \le \varepsilon_{i_1},..., p_{i_k}(a_{i_k}, b_{i_k}) \le \varepsilon_{i_k}$, виконано не менше q_1 та не виконано не більше q_2 нерівностей, де q_1, q_2 фіксовані додатні числа та вектор $(\varepsilon_1,...,\varepsilon_n)-$ де $\varepsilon_i \le 0$, $i = 1,...,n$, в протилежному випадку $B(w,S,S') = 0$.

Зрозуміло, що прямі обчислення за даною формулою обчислення оцінок будуть не ефективні, якщо взагалі практично можуть бути здійснені. В даний час запропоновано ряд підходів, що забезпечують більш ефективне обчислення оцінок та інші, що залежать від вибору системи опорних множин.

Зауважимо, що для часткового подолання цих обмежень введена модифікація поняття опорної множини (T-опорної множини), що не зменшує загальності даного поняття. Пропонується підхід, що забезпечує ефективний вибір таких опорних множин, при яких обчислення оцінок належності об’єкта S до класу з номером j значно спрощується. Приводяться дослідження по розширенню даного класу алгоритмів.

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

Мета роботи та задачі дослідження. Отже зважаючи на все вище сказане, метою даною роботи буде отримання ефективного способу побудови логічних дерев класифікації за допомогою поняття *T*-опорної множини та введення концепції алгоритмічного дерева класифікації. Саме це дозволить забезпечити простий та ефективний процес розпізнавання образів, а отже отримати найбільш адекватну та економічну форму результуючої схеми розпізнавання, що дозволить економити процесорний час на її роботу та оперативну пам‘ять для зберігання. Для досягнення даної мети були поставлені такі завдання:

- Показати особливості *T*-опорних множин в задачах розпізнавання образів.
- Дослідити питання зв‘язку *T*-опорних множин та логічних дерев класифікації, розширити моделі алгоритмів розпізнавання, які засновані на концепції ЛДК.
- Представити концепцію ЛДК в задачах розпізнавання шляхом введення дерев моделей класифікації та розпізнавання.

Питання *T*-опорних множин та деякі способи їх завдання. Перейдемо безпосередньо до головної ідеї *T*-опорної множини, яка полягає у відборі та фіксації деякого набору ознак разом зі своїми значеннями.

Визначення 1. *T*-опорна множина – це фіксований набір ознак з фіксованими їх значеннями:

$$\begin{pmatrix} e_{i_1} & & e_{i_k} \\ & ,..., & \\ X_{i_1} & & X_{i_k} \end{pmatrix}, \quad k = 1, ..., n .$$

Систему *T*-опорних множин будемо позначати Ω^T.

Нехай зафіксовані деякі числа q_1, q_2 та вектор $(\varepsilon_1, ..., \varepsilon_n)$, де $\varepsilon_i \geq 0,\ (i = 1, ..., n)$.

Визначення 2. *T*-опорна множина $\begin{pmatrix} e_{i_1} & & e_{i_k} \\ & ,..., & \\ X_{i_1} & & X_{i_k} \end{pmatrix}$ називається інформативною по відношенню до об‘єкта S, якщо в системі нерівностей $p_{i_1}(a_{i_1}, e_{i_1}) \leq \varepsilon_{i_1}, ..., p_{i_k}(a_{i_k}, e_{i_k}) \leq \varepsilon_{i_k}$ виконано не менше q_1 та не виконано не більше q_2 нерівностей.

Визначення 3. *T*-опорна множина $\begin{pmatrix} e_{i_1} & & e_{i_k} \\ & ,..., & \\ X_{i_1} & & X_{i_k} \end{pmatrix}$ називається інформативною по відношенню до класу j за інформацією $I(l)$, якщо в

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

системі нерівностей $p_{i_1}(a_{i_1}, e_{i_1}) \le \varepsilon_{i_1}, ..., p_{i_k}(a_{i_k}, e_{i_k}) \le \varepsilon_{i_k}$ виконано не менше q_1 та не виконано не більше q_2 нерівностей, для деяких об‘єктів тільки j-го класу. Відносно об‘єктів, що не входять в $I(l)$, ніяких обмежень не накладається.

Введемо наступні позначення – $w^T = \{w_1, ..., w_\xi\}$, де w_i буде i-ю T-опорною множиною, ξ – кількість T-опорних множин, H_j – множина об’єктів S, що належать класу H_j. Той факт, що T-опорна множина w інформативна до класу H_j, будемо позначати так: $w * H_j$, аналогічно $w * S$ – інформативність w до об‘єкту S. Запис $w * H_j$ означає, що існує хоча би один об‘єкт $S \in H_j$, що $w * H_j$.

<u>*Визначення 4.*</u> Систему T-опорних множин $\Omega^T = \{w_1, ..., w_\xi\}$, будемо називати інформативною, якщо:

$$1)\, \forall i \exists j (w_i * H_j);\ 2)\, \forall j \exists i (w_i * H_j);\ 3)\, \forall S \exists i (w_i * S);\ 4)\, \forall S^` \exists i (w_i * S^`).$$

Зауважимо, що тут $i = 1, ..., \xi;\ \ j = 1, ..., l$.

<u>*Визначення 5.*</u> Систему T-опорних множин $\Omega^T = \{w_1, ..., w_\xi\}$, будемо називати несуперечливою, якщо будуть виконуватись наступні умови:

$$1)\ \forall i \exists j (w_i * H_j);\ 2)\, \forall S^` \in H_j\ \exists i (w_i * S^`);$$

$$3)\, \bar{\exists} (i_1, i_2), i_1 \ne i_2, (w_{i_1} * S\ ma\ w_{i_n} * S) \Rightarrow (w_{i_1} * H_{j_1}\ ma\ w_{i_n} * H_{j_n}), i_1 \ne i.$$

Зауважимо, що тут $i = 1, ..., \xi;\ \ j = 1, ..., l$.

<u>*Визначення 6.*</u> Інформативну несуперечливу систему T-опорних множин $\Omega^T = \{w_1, ..., w_\xi\}$, назвемо простою, якщо буде виконуватися умова, що $\forall S \exists! i (w_i * S)$.

<u>*Завдання систем Т-опорних множин.*</u> На наступному етапі розглянемо способи завдання систем T-опорних множин $\Omega^T = \{w_1, ..., w_\xi\}$.

1) Найбільш природній спосіб завдання систем T-опорних множин $\Omega^T = \{w_1, ..., w_\xi\}$ – полягає у використанні їх основного визначення. Наприклад, випадковим чином фіксуються певні під-набори ознак та їх значення. Таким чином, використовуючи програмний генератор псевдо-випадкових чисел, можна сформулювати різні множини систем T-опорних множин. Після цього можна провести процедуру селекції і

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

виділити і інформативні, і несуперечливі, і інформативно-несуперечливі системи *T*-опорних множин $\Omega^T = \{w_1,..., w_\xi\}$.

2) Завдання систем *T*-опорних множин $\Omega^T = \{w_1,..., w_\xi\}$ за допомогою тестів. Нехай ознаки з номерами $i_1, ..., i_k$ утворюють тест, знайдений за деякою початковою інформацією $I(l)$. Якщо зафіксувати значення ознак, що входять в даний тест (звичайно, що з відповідними їх значеннями на $I(l)$), то дана множина буде утворювати несуперечливу систему *T*-опорних множин $\Omega^T = \{w_1,..., w_\xi\}$, але не завжди інформативну (знову же відповідно до визначення *T*-опорної множини). Відмітимо також, що відносно множини тестів, за якими зафіксована (побудована) система *T*-опорних множин $\Omega^T = \{w_1,..., w_\xi\}$, то тут ситуація аналогічна.

3) Зауважимо, що є можливість знаходити системи *T*-опорних множин $\Omega^T = \{w_1,..., w_\xi\}$, в тому числі інформативні, несуперечливі, прості використовуючи інші методи та алгоритми [8].

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

5) Знаходження простих систем *T*-опорних множин $\Omega^T = \{w_1,..., w_\xi\}$ за допомогою $I(l)$ та дерев розпізнавання має найбільше перспектив, оскільки мало залежить від об‘єму та виду початкової інформації $I(l)$, тобто ЛДК в певному сенсі є стисненим структурним типом даних. Наприклад:

- Інформація $I(l)$ може бути повністю розміщена в оперативній пам‘яті комп‘ютера.
- Інформація $I(l)$ представляє собою масив великого (надвеликого) об’єму, тобто значно більше об’єму оперативної пам’яті комп‘ютера.
- Інформація $I(l)$ представляє собою данні, в якій деякі значення ознак на деяких наборах об‘єктів невизначені.
- Набори деяких ознак в початковій інформації $I(l)$ є різнотипними.

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

Більш того, за допомогою ЛДК можна проводити деякі операції над системами *T*-опорних множин $\Omega^T = \{w_1,..., w_\xi\}$ та в результаті отримувати деякі інші прості системи *T*-опорних множин $\Omega^T = \{w_1,..., w_\xi\}$. Це дозволить (при виконанні деяких умов) знаходити базис *T*-опорних множин $\Omega^T = \{w_1,..., w_\xi\}$ та визначати всі інші набори множин через цей базис.

Питання зв'язку ЛДК та *T*-опорних множин. Розширення моделі алгоритмів розпізнавання, які засновані на концепції ЛДК.

Розглянемо в даній частині роботи деякі прості можливості застосування логічних дерев для вирішення поставлених задач розпізнавання образів. Нехай маємо бінарний випадок, тобто $M_i \in \{0,1\}$, $i = 1, ..., n$.

<u>*Визначення 7.*</u> ЛДК – це зв'язаний граф без циклів, в некінцевих вершинах якого знаходяться ознаки (узагальнені ознаки, алгоритми, узагальнені алгоритми), а ребра нумеруються значеннями цих ознак. В кінцевих вершинах (кінцевого ярусу) ЛДК знаходяться значення функції, що задає розбиття початкової інформації на класи або символ Δ невизначеності. При чому будуть виконуватись наступні умови:

а) На фіксованому шляху в ЛДК одна і таж сама ознака зустрічається тільки один раз (випадок регулярного дерева).

б) Об'єктам із $I(l)$ відповідає значення функції, що задає розбиття, при умові початкової відмінності між об'єктами різних класів (виконання гіпотези несуперечливості).

в) Довільному об'єкту $I(l)$ відповідає одно із значень множини $\{0, 1, ..., l\}$.

В даний час існують різні алгоритми та методи побудови ЛДК, що задовольняють даному визначенню та інші [9].

<u>*Визначення 8.*</u> Правилом розпізнавання (класифікації) являється деякий оператор Π, що приводить довільний допустимий об'єкт S за його інформаційним описом $I(S)$ в кінцеву вершину ЛДК (де знаходиться значення функції розпізнавання для даного об'єкту).

<u>*Визначення 9.*</u> Деякий алгоритм T називається алгоритмом розпізнавання (класифікації), якщо він переводить початкову інформацію $I(l)$ та інформаційний опис довільного числа q допустимих об'єктів $I(q)$ у вектор a_{r+q}, складений з елементів $\{0, 1, ..., l\}$ де $r = |I(l)|$ – кількість об'єктів, що складають початкову інформацію $I(l)$ при чому

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

об'єктам із $I(l)$ відповідає значення функції, що задає розбиття, при умові початкової відмінності між об'єктами різних класів.

Очевидно, що фіксоване ЛДК, побудоване за початковою інформацією $I(l)$ із заданим правилом розпізнавання (класифікації) задає деякий алгоритм розпізнавання.

Теорема 1. Довільне ЛДК (за даними деякої початкової НВ) задає систему T-опорних множин, причому кількість T-опорних множин в цій системі дорівнює кількості кінцевих вершин ЛДК, а кількість інформативних T-опорних множин до заданих класів дорівнює кількості визначених вершин ЛДК (рис. 1). Для даного прикладу загальна кількість T-опорних множин складає 8, причому серед них будуть інформативні до класу H_1 – дві, інформативні до класу H_2 – три, інформативні до класу H_3 – три.

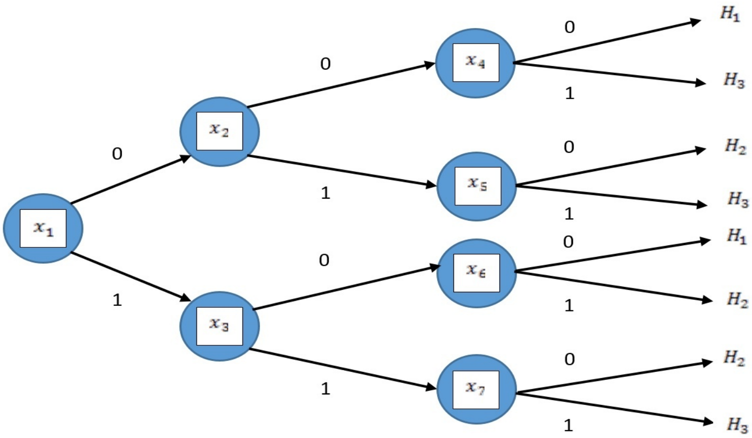

Рис. 1. Приклад ЛДК, яке задає вісім T-опорних множин

Зауваження. Під визначеною вершиною ЛДК, розуміємо вершину в якій знаходиться конкретна мітка зі значенням функції розпізнавання (число або константа).

Тобто визначена вершина зазвичай фіналізує деякий шлях (схему розпізнавання) в ЛДК.

Наслідок. Якщо в ЛДК всі кінцеві вершини визначені, то воно задає просту систему T-опорних множин.

Доведення. Зафіксуємо в ЛДК деяку його визначену вершину. Зрозуміло, що цій вершині відповідає цілком визначений шлях в ЛДК, що зв'язує її із початковою вершину ЛДК та характеризується деяким набором ознак і їх фіксованих значень. Це ніщо інше, як T-опорна множина. Проробимо аналогічну процедуру зі всіма кінцевими

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

вершинами ЛДК. Очевидно, що кожна кінцева вершина в ЛДК задає деяку *T*-опорну множину. Якщо кінцева вершина ЛДК визначена, то *T*-опорна множина, відповідна цій вершині, буде інформативною до класу, номер якого знаходиться в цій вершині, так як об‘єкти що попадають в цю вершину, належать одному фіксованому класу.

Доведення наслідка. Оскільки всі кінцеві вершини ЛДК визначені, то:

1) Дерево визначає тільки інформативні *T*-опорні множини до відповідних класів.

2) Кожний клас в ЛДК характеризується, щонайменше, одною кінцевою вершиною, яка в свою чергу визначає інформативну *T*-опорну множину до цього класу.

3) Довільний допустимий об’єкт класифікації S за допомогою деякого правила розпізнавання (класифікації) приводиться в одну з кінцевих вершин ЛДК, але ця кінцева вершина визначає інформативну *T*-опорну множину по відношенню до нього (рис. 2).

Визначимо формально алгоритм побудови деякого ЛДК. Візьмемо за основу, наприклад, алгоритм, описаний в [10]. Як правило, всі відомі алгоритми побудови ЛДК засновані на цій ідеї. Відмінність їх полягає у виборі оцінки інформативності ознак на відповідному кроці побудови ЛДК. Якщо опустити оцінку інформативності ознак при побудові ЛДК, то можна ввести поняття випадкового дерева класифікації (ВДК).

Визначення 10. ВДК називається ЛДК, ознаки в вершинах якого у процесі його побудови вибираються випадковим чином (програмним генератором PRG).

Будемо розглядати тільки такі ВДК, на фіксованому шляху яких одна і та же ознака зустрічається тільки один раз (ВДК першого типу). Очевидно, що довільне ЛДК можна звести до такого без втрати інформації. Множина алгоритмів, що визначаються простими системами *T*-опорних множин, що задаються ВДК, являється неповною. Цей факт легко доводиться контр-прикладом.

Для розширення цієї моделі введемо поняття оператора над ВДК.

Нехай $D = \{D_1, ..., D_\xi\}$ – деяка фіксована множина ВДК, що побудована для фіксованої задачі розпізнавання (наприклад за алгоритмом з роботи [11]). Нехай $D_i \in D$ та C – довільна його вершина, що однозначно відповідає його піддереву (фіксує в його структурі) $D_i(C)$ та $D_j \in D$, $j \neq i$ (рис. 2). Введемо операцію $\coprod$ над D_i та D_j, яка полягає у заміні $D_i(C)$ в D_i деревом розпізнавання D_j, тобто $D_i^j = (D_i(C))\coprod D_j$.

<u>*Теорема 2.*</u> $D_i^j = (D_i(C))\coprod D_j$ – являється деревом розпізнавання, $i, j = 1, ..., \xi$, де ξ – загальна кількість ЛДК (або ВДК), C – довільна вершина ЛДК D_i.

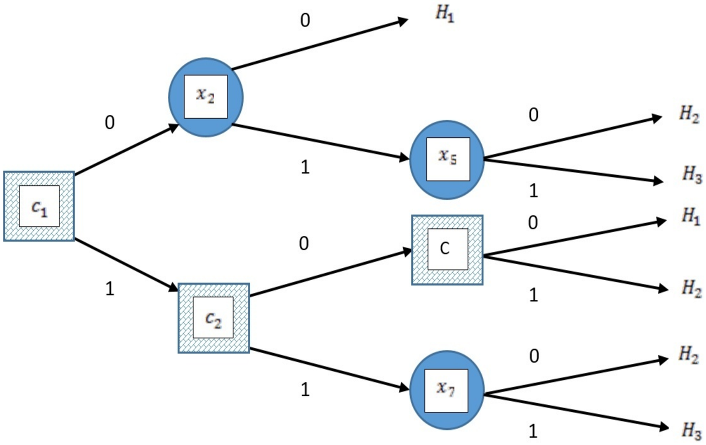

Рис. 2. Виділення вершиною C піддерева $D_i(C)$ в структурі ЛДК D_i

<u>*Визначення 11.*</u> Нехай D деяка множина ВДК. Замиканням $[D]$ множини D називається сукупність всіх ЛДК з P_k, які одержані в результаті дії оператора підстановки над ВДК з D.

Операція отримання $[D]$ з D називається операцією замикання.

<u>*Визначення 12.*</u> Деяка множина ВДК – P, для яких виконується умова $[P] = D$ називається неприводимою системою, якщо замикання любої власної підмножини P' з P, $(P' \subset P)$ відмінно від замикання всієї підмножини P, тобто $[P'] \subset [P]$, $[P'] \neq [P]$.

<u>*Визначення 13.*</u> Неприводима, повна в замкнутому класі система ВДК називається базисом множини ВДК – D.

<u>*Визначення 14.*</u> Правило розпізнавання, класифікації (оператор Π) задіює ознаки $P_{i_1}, ..., P_{i_\xi}$ (у вигляді вершин ЛДК) за допомогою інформації $I(S)$, якщо вони зустрічаються на шляху, що приводить цей об‘єкт в кінцеву вершину ЛДК.

Позначимо цей факт наступним чином $\Pi(I(S)) = \{P_{i_1}, ..., P_{i_\xi}\}$. Запис $D_i(I(S)) = j$, означає віднесення об‘єкта S до класу j за допомогою

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

деякого ЛДК – D_i. Через Π_i позначимо правило розпізнавання, класифікації (оператор) для фіксованого ЛДК D_i.

Теорема 3. Множина $D = \{D_1, ..., D_\xi\}$ утворює базис задачі розпізнавання, якщо виконуються наступні умови (визначеності та повноти):

1) $\forall S \in I(q), \forall j,\ j = 1,...,l;\ \exists i,\ i = 1,...,\xi;\ D_i(I(S)) = j$;
2) $\Pi_1(I(S)) \cup ... \cup \Pi_\xi(I(S)) = \{P_1,...,P_n\}$.

Зауважимо, що тут n – загальна кількість ознак задачі розпізнавання (в інформаційному описі об'єкту НВ), S – об'єкт з множини допустимих об'єктів $I(q)$, $D_i(I(S)) = j$ – віднесення деякого об'єкта S до класу j за допомогою ЛДК D_i та його інформаційним описом $I(S)$.

Питання особливостей дерев моделей класифікації та розпізнавання. На даному етапі піднімемо важливе питання – особливостей дерев моделей класифікації та розпізнавання (ДМК).

Нехай маємо $\eta_k = \{i_1, ..., i_k\},\ k \in \{1, ..., n\},\ I(S_j^\eta) = \{b_{i_1}^j, ..., b_{i_k}^j\}$ – частковий опис деякого допустимого об'єкта $S_j, b_i^j \in M_i$, тут M_i – ознакова область (метричний простір з метрикою p_i, $i = 1, ..., n$, P_{η_k} – деякий k-ий предикат $P_{\eta_k}(b_{i_1}^j, ..., b_{i_k}^j)$, заданий на множині часткових (ознакових) описів допустимих об'єктів S_j, $\{I(_j^{\eta_k}), P_{\eta_k}\}$ – множина моделей, що представляють деякі алгебраїчні системи розпізнавання у вигляді логічних дерев. При фіксованому η_k та P_{η_k} будемо мати деяку конкретну модель $\{I(_j^{\eta_k}),\ P_{\eta_k}\}$.

Визначення 15. По аналогії з логічним деревом, дерево моделей (ДМ) – це зв'язаний граф без циклів, в некінцевих вершинах якого знаходяться фіксовані моделі з $\{I(_j^{\eta_k}),\ P_{\eta_k}\}$, а ребра нумеруються значеннями предикатів з цих моделей. В кінцевих вершинах розташовані значення функції, що задає розбиття $I(l)$ на класи або символ Δ невизначеності. Причому на фіксованому шляху ДМ одна й таж сама модель може зустрічатися тільки один раз.

Визначення 16. Аналогічно, правилом розпізнавання (класифікації) ДМ являється деякий оператор, що відносить довільний допустимий

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

об‘єкт S, (за його початковим інформаційним описом $I(S)$) в фіксовану кінцеву вершину дерева моделей.

Визначення 17. Аналогічно, дерево моделей класифікації та розпізнавання (ДМК) – це деяке ДМ із заданим правилом розпізнавання (класифікації), для якого об’єктам з $I(l)$ відповідає значення функції розпізнавання, що задає розбиття $I(l)$ на класи.

Очевидно, що довільне ДМК – являється алгоритмом розпізнавання та класифікації (у вигляді логічного дерева) деяких об‘єктів за початковою інформацією НВ.

Отже, відповідно до приведеного визначення ДМК можна запропонувати наступні означення інформативності по відношенню до об‘єкту S та інформативності для деякого класу j за початковою інформацією $I(l)$.

Визначення 18. Деяка модель $\{I(^{\eta_k}_j),\ P_{\eta_k}\}$ називається інформативною по відношенню до об‘єкту S, якщо значення предиката P_{η_k} вірно на його частковому (ознаковому) описі.

Визначення 19. Модель $\{I(^{\eta_k}_j),\ P_{\eta_k}\}$ називається інформативною для деякого класу j за початковою інформацією $I(l)$, якщо вона інформативна тільки для деяких об‘єктів цього класу. Відносно об‘єктів, що не входять в $I(l)$, ніяких обмежень не накладається.

Визначення 20. Просте ДМК – це ДМК, на кожному ярусі якого, крім початкового, знаходяться по дві вершини.

Теорема 4. ДМК, в вершинах якого знаходяться моделі інформативні по відношенню до відповідних класів – є простим.

Нехай $V = \{V_1, ..., V_\xi\}$ – деяка множина ДМК, побудованих для фіксованої задачі. Введемо наступний оператор підстановки над V. Нехай $V_i \in V$ та D_j полягає в заміні $D_i(C)$ в D_j ДМК $D_j \to V_i^j = (V_i, C) \cup V_j$.

Теорема 5. $V_i^j = (V_i, C) \cup V_j$ представляє собою дерево.

Ця теорема дозволяє ввести операцію замикання над ДМК, таку же саму як над ЛДК та дослідити можливість знаходження базису задачі над множиною ДМК.

Відмітимо також принциповий момент, що в якості $\{I(^{\eta_k}_j),\ P_{\eta_k}\}$ (тобто вершин логічного дерева) можна взяти любі відомі алгоритми розпізнавання (отже, в якісь мірі, ми вперше прийшли до концепції АДК – в вершинах якого знаходяться окремі, автономні алгоритми

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

розпізнавання та класифікації). Тоді можна зафіксувати, що таке ДМК буде представляти собою орієнтований граф без циклів (по аналогії з ЛДК), в некінцевих вершинах якого знаходяться відомі алгоритми розпізнавання (вершини логічного дерева є окремими алгоритмами), стрілки нумеруються значеннями алгоритму на об‘єктах, а в кінцевих вершинах знаходяться деякі значення функції розпізнавання, що задає початкове розбиття інформації $I(l)$ на класи.

Зважаючи на все вище сказане можна зафіксувати наступні пункти:

1) Головною ідеєю *T*-опорної множини, є відбір (фіксація) певного набору ознак разом зі своїми значеннями на основі інформації деякої початкової НВ, з можливістю наступної оцінки даних опорних множин за допомогою певних функціоналів [8].

2) Відмітимо, що існує багато способів завдання систем *T*-опорних множин $\Omega^T = \{w_1, ..., w_\xi\}$, причому самий простий і природній – випливає безпосередньо з самого визначення даного поняття і полягає в фіксації деякого набору ознак та їх значень на основі програмного генератора PRG.

3) Зауважимо, що процес знаходження простих систем *T*-опорних множин $\Omega^T = \{w_1, ..., w_\xi\}$ за допомогою масиву деякої початкової інформації та дерев розпізнавання мало залежить від об‘єму, структури та виду початкової інформації $I(l)$, тобто побудова логічного дерева представляє собою процес стиснення початкових даних НВ.

4) На сьогоднішній день існують різноманітні методи побудови як ЛДК з одноразовим використанням ознак в структурі логічного дерева (алгоритми випадкових дерев, метод розгалуженого вбору ознак з початковою оцінкою інформативності), так і ЛДК з повторами різних ознак на ярусах логічного дерева (алгоритм побудови ЛДК з покроковою оцінкою важливості ознак).

5) Деяке правило розпізнавання (класифікації) приводить довільний допустимий об‘єкт S за його інформаційним описом $I(S)$ в кінцеву вершину ЛДК, тобто в деякій мірі задає фіксований шлях в структурі логічного дерева.

6) Зауважимо, що якщо зафіксувати деяку вершину C в структурі довільного ЛДК – D, то вона буде однозначно виділяти певне піддерево D^* (фіксувати в його структурі).

7) Зауважимо, що довільне правило розпізнавання, класифікації задіює ознаки $P_{i_1}, ..., P_{i_\xi}$ (у вигляді вершин ЛДК) за допомогою інформації $I(S)$, якщо вони зустрічаються на шляху, що приводить цей об‘єкт в кінцеву вершину ЛДК.

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

8) Деяка множина моделей $\{I(^{\eta_k}_j),\ P_{\eta_k}\}$, що представляють алгебраїчні системи розпізнавання (у вигляді структур логічних дерев) при фіксованому η_k та P_{η_k} перетворюється на деяку конкретну модель $\{I(^{\eta_k}_j),\ P_{\eta_k}\}$ (деревоподібну логічну структуру).

9) Дерево моделей представляє собою зв‘язаний граф без циклів, в некінцевих вершинах якого знаходяться фіксовані моделі з $\{I(^{\eta_k}_j),\ P_{\eta_k}\}$, а ребра нумеруються значеннями предикатів з цих моделей, причому в кінцевих вершинах знаходяться значення функції, що задає розбиття $I(l)$ на класи або символ Δ невизначеності, а на фіксованому шляху ДМ одна й таж сама модель може зустрічатися тільки один раз.

10) Відмітимо, що правилом розпізнавання (класифікації) ДМ являється деякий оператор, що відносить довільний допустимий об‘єкт S, (за його початковим інформаційним описом) в фіксовану кінцеву вершину дерева моделей.

11) Отже ДМК – це деяке ДМ із заданим правилом розпізнавання (класифікації), для якого об’єктам з НВ відповідає значення ФР, що задає розбиття початкової вибірки на класи.

12) Відмітимо, що просте ДМК – це ДМК, на кожному ярусі якого, крім початкового, знаходяться по дві вершини, а ДМК, в вершинах якого знаходяться моделі інформативні по відношенню до відповідних класів – є простим.

13) Відмітимо також, що ДМК в вершинах якого знаходяться моделі інформативні по відношенню до відповідних класів – є простим ДМК.

Висновки. Отже в даному дослідженні були виконані наступі завдання:

- Запропонована загальна ідея T-опорних множин в розрізі концепції логічних дерев класифікації, що дозволило зв‘язати методи обчислення оцінок та ЛДК. Показані основні способи представлення T-опорних множин в задачах розпізнавання образів (в розрізі концепції ЛДК), продемонстроване графічне представлення такої опорної множини в структурі логічного дерева. Такий підхід дозволив визначити інформативність щодо класу та об‘єкту класифікації відносно структури логічного дерева.
- Досліджене питання зв‘язку T-опорних множин та логічних дерев класифікації, показаний зв‘язок шляху в структурі логічного дерева (фіксованого правила класифікації) та T-опорної множини. Такий підхід дозволяє забезпечити ефективний пошук методів оптимізації структури

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

результуючого ЛДК.

- Представлена концепцію алгоритмічного дерева класифікації (в вершинах якого знаходяться окремі, автономні алгоритми розпізнавання та класифікації) в задачах розпізнавання шляхом введення дерев моделей класифікації. Відмітимо, що такий підхід забезпечує високу універсальність побудованих моделей розпізнавання та дозволяє раціонально використовувати накопичений потенціал методів та алгоритмів класифікації.

Список літератури:

1. *Zheng Z.* Real world performance of association rule algorithms / *Z. Zheng, R. Kohavi, L. Mason* // 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.* Концептуальна основа систем розпізнавання образів на основі метода розгалуженого вибору ознак / *Е.Ю. Василенко, І.Ф. Повхан, Ф.Г. Ващук* // Науково технічний журнал "European Journal of Enterprise Technologies". – 2004. – № 7 [1]. – С. 13-15.

3. *Василенко Ю.А.* Метод розгалуженого вибору ознак в математичному конструюванні багаторівневих систем розпізнавання образів / *Е.Ю. Василенко, І.Ф. Повхан, Ф.Г. Ващук* // Науково технічний журнал "Штучний Інтелект". – 2003. – № 7. – С. 246-249.

4. *Василенко Ю.А.* Проблема оцінки складності логічних дерев розпізнавання та загальний метод їх оптимізації / *Ф.Г. Ващук, Ю.А. Василенко, І.Ф. Повхан* // Науково технічний журнал "European Journal of Enterprise Technologies". – 2011. – 6/4 (54). – С. 24-28.

5. *Василенко Ю.А.* Загальна оцінка мінімізації деревоподібних логічних структур / *Ф.Г. Ващук, І.Ф. Повхан* // Науково технічний журнал "European Journal of Enterprise Technologies". – 2012. – 1/4 (55). – С. 29-33.

6. *Quinlan J.R.* Induction of Decision Trees / *J.R. Quinlan* // Machine Learning. – 2008. − № 1. − P. 1-81.

7. *Vtogoff P.E.* Incremental Induction of Decision Trees / *P.E. Vtogoff* // Machine Learning. − 2009. − № 4. − P. 161-186.

8. *Повхан І.Ф.* Проблема функціональної оцінки навчальної вибірки в задачах розпізнавання дискретних об‘єктів / *І.Ф. Повхан* // Вчені записки Таврійського національного університету. Серія: технічні науки. – 2018. – Том 29 (68). – № 6. – С. 217-222

9. *Povhan I.* Designing of recognition system of discrete objects / *I. Povhan* // 2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP), Lviv. – 2016, Ukraine. – P. 226-231.

10. *Лавер В.О.* Алгоритми побудови логічних дерев класифікації в задачах розпізнавання образів / *І.Ф. Повхан* // Вчені записки Таврійського національного університету. Серія: технічні науки. – 2019. – Том 30 (69). – № 4. – С. 100-106.

11. *Повхан І.Ф.* Особливості випадкових логічних дерев класифікації в задачах розпізнавання образів / *І.Ф. Повхан* // Вчені записки Таврійського національного університету. Серія: технічні науки. – 2019. – Том 30 (70). – № 5. – С. 156-166.

12. *Srikant R.* Mining generalized association rules / *R. Srikant, R. Agrawal* // Future

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

Generation Computer
13. *Toivonen H.* Sampling large databases for association rules / *H. Toivonen* // In Proc. 1996 Int. Conf. Very Large Data Bases / *Ed. by T. M. Vijayaraman, A. P. Buchmann, C. Mohan*, *N. L. Sarda*. — Morgan Kaufman, 1996. – P. 134-145.
14. *Whitley D.* An overview of evolutionary algorithms: practical issues and common pitfalls / *D. Whitley* // Information and Software Technology. – 2001. – Vol. 43. – № 14. – P. 817-831.
15. *Bodyanskiy Y.* Hybrid neuro-neo-fuzzy system and its adaptive learning algorithm / *Y. Bodyanskiy, O. Vynokurova, G. Setlak and I. Pliss* // Xth Scien. and Tech. Conf. "Computer Sciences and Information Technologies" (CSIT). – 2015. – Lviv. – P. 111-114.

References:

1. Zheng, Z., Kohavi, R., and Mason, L. (2001), "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*, pp. 401-406.
2. Vasilenko, Y.A., Povhan, I.F., and Vaschuk, F.G. (2004), "Conceptual basis of pattern recognition systems based on the method of branched feature selection", *Scientific and technical journal "European Journal of Enterprise Technologies"*, 7 [1], pp. 13-15.
3. Vasilenko, Y.A., Povhan, I.F., and Vaschuk, F.G. (2003), "Method of branched feature selection in mathematical design of multilevel pattern recognition systems", *Scientific and technical journal "Artificial Intelligence",* No. 7, pp. 246-249.
4. Vasilenko, Y.A., Povhan, I.F., and Vaschuk, F.G. (2011), "The problem of evaluation of complexity of logic trees, recognition, and a general method of optimization", *Scientific and technical journal "European Journal of Enterprise Technologies"*, 6/4 (54), pp. 24-28.
5. Vasilenko, Y.A., Povhan, I.F., and Vaschuk, F.G. (2012), "Overall assessment of minimization of logical tree structures", *Scientific and technical journal "European Journal of Enterprise Technologies"*, 1/4 (55), pp. 29-33.
6. Quinlan, J.R. (2008), "Induction of Decision Trees", *Machine Learning,* No. 1, pp. 1-81.
7. Vtogoff, P.E. (2009), "Incremental Induction of Decision Trees", *Machine Learning,* No. 4, pp. 161-186.
8. Povkhan, I.F. (2018), "The problem of functional evaluation of the training sample in the problems of recognition of discrete objects", *Scientific notes of Taurida national University. Series: technical Sciences,* Vol. 29 (68), No. 6, pp. 217-222.
9. Povkhan, I. (2016), "Designing of recognition system of discrete objects", *2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP)*, Lviv, Ukraine, pp. 226-231.
10. Laver, V.A., and Povhan, I.F. (2019), "Algorithms for constructing logical classification trees in pattern recognition problems", *Scientific notes of Taurida national University. Series: technical Sciences,* Vol. 30 (69), No. 4, pp. 100-106.
11. Povkhan, I.F. (2019), "Features of random logical classification trees in pattern recognition problems", *Scientific notes of Tauride national University. Series: technical Sciences,* Vol. 30 (70), No. 5, pp. 156-166.
12. Srikant, R., and Agrawal, R. (2015), "Agrawal R. Mining generalized association rules", *Future Generation Computer Systems,* Vol. 13, No. 2, pp. 161-180.
13. Toivonen, H. (1996), "Sampling large databases for association rules", *In Proc. 1996 Int. Conf. Very Large Data Bases* / Ed. by T.M. Vijayaraman, A.P. Buchmann, C. Mohan,

N.L. Sarda, Morgan Kaufman, pp. 134-145.
14. Whitley, D. (2001), "An overview of evolutionary algorithms: practical issues and common pitfalls", *Information and Software Technology,* Vol. 43, No. 14, pp. 817-831.
15. Bodyanskiy Y., Vynokurova O., Setlak G., and 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, pp. 111-114.

Статтю представив д.т.н, проф. Головач Й.Г. (*кафедра програмного забезпечення систем ДВНЗ Ужгородського національного університету*).

Надійшла/Received: 12.09.2019

Povhan Igor, Cand. Tech. Sci.
Uzhgorod national university, Ukraine
Narodna Square 3, Uzhgorod, Ukraine, 88000
e-mail: Igor.povkhan@uzhnu.edu.ua
ORCID ID: 0000-0002-7034-8702

УДК 004.8: 004.89: 519.7

Повхан І.Ф. Питання побудови деревоподібних моделей розпізнавання образів / І.Ф. Повхан // Вісник НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2019. – № 28 (1353). – С. 39 – 57.

Робота піднімає важливі питання теорії розпізнавання образів (дискретних об‘єктів), які пов‘язані з загальною проблематикою побудови деревоподібних схем розпізнавання, класифікації та плавно підводить до концепції алгоритмічного дерева класифікації, що безумовно є найвищим рівнем абстракції в теорії дерев розпізнавання. Простий, ефективний, економний метод побудови логічного дерева класифікації навчальної вибірки дозволяє забезпечити необхідну швидкодію, рівень складності схеми розпізнавання, що гарантує проведення простого та повного розпізнавання дискретних об‘єктів. На сьогоднішній день існують різноманітні методи побудови як логічних дерев з одноразовим використанням ознак в структурі логічного дерева (алгоритми випадкових дерев, метод розгалуженого вбору ознак з початковою оцінкою інформативності), так і дерев з повторами різних ознак на ярусах логічного дерева (алгоритм побудови дерева з покроковою оцінкою важливості ознак, тощо). В роботі фіксуються суттєві переваги логічних дерев класифікації – програмна простота побудови дерева класифікації, зменшення часу загальної генерації логічного дерева та інше. Іл.: 2. Бібліогр.: 15 назв.

Ключові слова: теорія розпізнавання образів; логічне дерево; алгоритмічні дерева класифікації; навчальна вибірка.

УДК 004.8: 004.89: 519.7

Повхан И.Ф. Вопрос построения древовидных моделей распознавания образов / И.Ф. Повхан // Вестник НТУ "ХПИ". Серия: Информатика и моделирование. – Харьков: НТУ "ХПИ". – 2019. – № 28 (1353). – С. 39 – 57.

Работа поднимает важные вопросы теории распознавания образов (дискретных объектов), которые связаны с общей проблематикой построения древовидных схем распознавания, классификации и плавно подводит к концепции алгоритмического дерева классификации, что безусловно является самым высоким уровнем абстракции в теории деревьев распознавания. Простой, эффективный, экономичный метод построения логического дерева классификации обучающей выборки позволяет обеспечить необходимое быстродействие, уровень сложности схемы распознавания, что гарантирует проведение простого и полного распознавания дискретных объектов. На сегодняшний день существуют различные методы построения как логических деревьев с однократным использованием признаков в структуре логического дерева (алгоритмы случайных деревьев, метод разветвленного выбора признаков с начальной оценкой информативности), так и деревьев с повторами различных признаков на ярусах логического дерева (алгоритм построения дерева с пошаговой оценкой важности признаков, и тому подобное). В работе фиксируются существенные преимущества логических деревьев классификации – программная простота построения дерева классификации, уменьшение времени общей генерации логического дерева и прочее. Ил.: 2. Библиогр.: 15 назв.

Ключевые слова: теория распознавания образов; логическое дерево; алгоритмические деревья классификации; обучающая выборка

UDC 004.8: 004.89: 519.7

Povhan I.F. The question of building tree models of pattern recognition / I.F. Povkhan // Herald of the National Technical University "KhPI". Series of "Informatics and Modeling". – Kharkov: NTU "KhPI". – 2019. – № 28 (1353). – P. 39 – 57.

The work raises important questions of the theory of pattern recognition (discrete objects), which are associated with the General problem of building tree-like recognition schemes, classification and smoothly leads to the concept of algorithmic classification tree, which is certainly the highest level of abstraction in the theory of recognition trees. It is clear that a simple, efficient, cost-effective method of constructing a logical tree of classification of the training sample allows you to provide the necessary speed, the level of complexity of the recognition scheme, which guarantees a simple and complete recognition of discrete objects. So, today, there are different construction methods as a logical tree with a single use of signs in the structure of the logical tree (the algorithms of random trees method extensive array of signs with an initial assessment of informational content), and trees with repetitions of various characteristics on the tier of the logical tree (algorithm of the tree building step-by-step assessment of the importance of signs, and the like). The paper fixes the significant advantages of logical classification trees-software simplicity of building a classification tree, reducing the time of General generation of a logical tree, and so on. Figs.: 2. Refs.: 15 titles.

Keywords: pattern recognition theory; logical tree; algorithmic classification trees; training sample.