

## БАГАТОВИМІРНА КЛАСИФІКАЦІЯ В ЗАДАЧАХ МОНІТОРИНГУ ПІДТОПЛЕНЬ ЗАСОБАМИ ДИСТАНЦІЙНОГО ЗОНДУВАННЯ ЗЕМЛІ

© М.Н. Жуков, А.М. Макаренко, 2009

*Київський національний університет імені Тараса Шевченка, Київ, Україна*

The article estimates the possibility of using the Bayes approach as compared to the SVM for monitoring underflooding problems. Ways of realization of the theoretically optimal criterion are considered. The conclusion about preference of the non-parametrical classification, which algorithm is described, is made. Techniques of enhancement of classification reliability are examined.

Підтоплення, згідно з проектом державних будівельних норм (ДБН) [1], – це інженерно-геологічний процес, який має прояви у певних умовах природного середовища внаслідок спільного впливу причин і факторів як природного, так і техногенного походження, коли за розрахунковий період часу підвищується вологість ґрунтів або піднімається рівень підземних вод до граничних значень, за якими порушуються умови будівництва й експлуатації будинків і споруд, відбуваються ушкодження та загибель зелених насаджень.

Вищеозначені природні та техногенні причини і фактори є складною системою, що призводить до виникнення підтоплень. Перевага тих чи інших факторів або їх співвідношення визначає вид підтоплення – стале (постійне) підняття рівня ґрунтових вод, періодичні підняття рівня ґрунтових вод, накопичення дощової (талого) води, підняття рівнів води, фільтрація з водойм і каналів, втрати з водонесівних мереж і комунікацій, зрошення земель тощо [1].

Дистанційне зондування забезпечує можливість моніторингу підтоплень завдяки залежності відбивної здатності води від довжини хвиль [2]. Ця залежність полягає у тому, що, чим більшою є довжина хвилі, тим менший коефіцієнт відбиття і, відповідно, інтенсивніше поглинання хвиль. Інфрарчервоні (ІЧ) хвилі майже повністю поглинаються водою. Тому за дистанційного зондування на знімках в ІЧ-діапазоні підтоплені ґрунти виглядатимуть тим темнішими, чим більше вони зволожені. Найтемнішими виглядають відкриті водні поверхні. Це і є основною ознакою для виділення підтоплень. Виділення зон підтоплень може бути якісним або кількісним. У другому випадку інтерпретація носить назву класифікації даних, або, за інтерпретації зображень, – розпізнавання образів.

Одним з поширених підходів до розв'язання задачі інтерпретації знімків є застосування методу багатовимірної класифікації. До вирішення цієї задачі класифікації існують підходи, які

умовно можна характеризувати як емпіричний та параметричний.

**1. Постановка задачі.** Космічний знімок є матрицею [3], кожен елемент якої (піксел) характеризується вектором ознак  $y_{ij} = \{y_1, y_2, \dots, y_m\}_{ij}$ . Компонентами вектора є значення інтенсивності відбитої або випромєненої енергії, зарєєстрованої в різних спектральних каналах. Кількість компонент вектора  $m$  визначається кількістю дискретних каналів, у яких проводиться зйомка;  $i, j$  – умовні координати піксела на знімку.

Кожен піксел характеризує певну ділянку земної поверхні. Розмір цієї ділянки визначає роздільну здатність знімка. Розрізняють знімки низької (500 м у пікселі та більше), середньої (500–100 м/піксел), високої (90–10 м/піксел) та надвисокої (1 м/піксел і менше) роздільної здатності. Залежно від масштабу поставленої задачі (якого розміру та інтенсивності підтоплення необхідно локалізувати) вибирають необхідну роздільну здатність. Для регіональних досліджень придатні знімки низької та середньої роздільної здатності, для локальних – високої. Оскільки найбільший інтерес становить використання дистанційного зондування Землі (ДЗЗ) для моніторингу локальних підтоплень, то для вирішення цієї задачі потрібні знімки високої та надвисокої роздільної здатності.

**Загальна постановка задачі.** Маємо  $t$  класів  $Q_j, j = \overline{1, t}$ , кожному з яких відповідає певний розподіл  $m$ -вимірної випадкової величини  $\xi$ . Компонентами її слугують значення інтенсивності відбитої або випромєненої енергії, зарєєстрованої в  $m$  спектральних каналах супутникової зйомки:  $\xi = (\xi_1, \xi_2, \dots, \xi_m)^T$ . Класифікації підлягає кожен із зазначених векторів  $y_{ij}$ , результатом якої є визначення класу або групи класів, до якого (якої) належить  $y_{ij}$ . Розширена постановка задачі передбачає паралельну оцінку достовірності класифікації кожного вектора  $y_{ij}$ .

**2. Емпіричний підхід (метод опорних векторів).**

**Випадок лінійно роздільних класів.** Характерний приклад емпіричного підходу – метод опор-

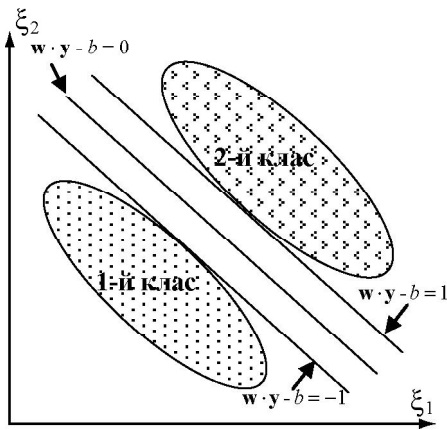


Рис. 1. Побудова лінійної роздільної площини для випадку лінійно роздільних класів

них векторів (SVM – support vector machines), запропонований В.Н. Вапніком [4]. Ґрунтується на знаходженні аналітичного вигляду оптимальних роздільних гіперплощин, що розмежовують сукупність об'єктів у  $m$ -вимірному просторі. Метод потребує “навчання”, зміст якого полягає у визначенні невідомих параметрів роздільної гіперплощини із використанням навчальних вибірок – сукупності об'єктів, належність яких до того чи іншого класу є наперед відомою.

Якщо класи можливо розділити гіперплощиною, то, вочевидь, вона буде не єдиною. Ідея методу SVM полягає у виборі з усієї сукупності можливих гіперплощин лише одної, сумарна відстань від якої до найближчих об'єктів різних класів буде максимальною.

Рівняння роздільної гіперплощини запишемо у вигляді

$$\mathbf{w} \cdot \mathbf{y} - b = 0, \quad (1)$$

де  $\mathbf{w}$  – вектор нормалі до гіперплощини (ваговий вектор);  $b$  – зміщення гіперплощини відносно початку координат.

Критерій класифікації за змістом аналогічний критерію максимізації відстані між двома гіперплощинами, паралельними оптимальній роздільній гіперплощині та побудованими на об'єктах різних класів навчальної множини. Такі об'єкти називають опорними векторами.

Після нормування ці паралельні гіперплощини можуть бути описані рівняннями (рис.1)

$$\begin{aligned} \mathbf{w} \cdot \mathbf{y} - b &= 1, \\ \mathbf{w} \cdot \mathbf{y} - b &= -1. \end{aligned} \quad (2)$$

Якщо навчальна множина є лінійно роздільною, то можна вибрати гіперплощини таким чином, щоб між ними не було жодного об'єкта, і потім максимізувати відстань між гіперплощинами. Ширина смуги дорівнює  $\frac{2}{\|\mathbf{w}\|}$ . Отже, задача побудови оптимальної роздільної гіперплощини на заданій навчальній множині зводиться до мінімізації  $\|\mathbf{w}\|$  за одночасного виконання умов

$$\begin{aligned} \mathbf{w} \cdot \mathbf{y}_i - b &\geq 1 \text{ для } \mathbf{y}_i = 1, \\ \mathbf{w} \cdot \mathbf{y}_i - b &\leq -1 \text{ для } \mathbf{y}_i = -1. \end{aligned} \quad (3)$$

Формалізована постановка задачі зводиться до задачі квадратичної оптимізації, що має вигляд

$$\begin{cases} \|\mathbf{w}\|^2 \rightarrow \min, \\ c_i (\mathbf{w} \cdot \mathbf{y}_i - b) \geq 1. \end{cases} \quad (4)$$

Зазвичай для розв'язання цієї задачі оптимізації використовують метод Лагранжа [5].

#### Випадок нелінійних роздільних гіперповерхонь.

На рис. 2. показано випадки, за яких класи неможливо розділити за допомогою лінійних площин. Метод SVM продовжує розвиватися у напрямку використання нелінійних класифікаторів, в основі яких лежить перехід від скалярних добутоків до довільних ядер (kernel function), якими залежно від задач можуть бути поліноміальні функції, радіальні функції Гауса тощо.

Результуючий алгоритм подібний до алгоритму лінійної класифікації з тією відмінністю, що кожен скалярний добуток у наведених вище формулах замінюємо на нелінійну функцію ядра. Таким чином, дані переводимо у простір вищої вимірності, в якому вже може існувати оптимальна роздільна гіперплощина. Відповідна їй функція у вихідному двовимірному просторі буде нелінійною (рис. 2.) Тестові приклади такої класифікації були реалізовані за допомогою інтерак-

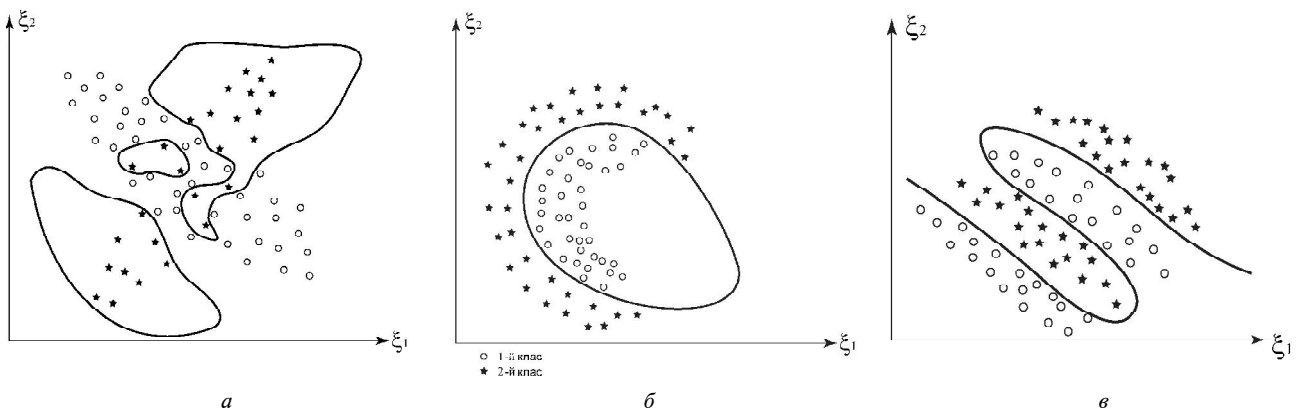


Рис. 2. Результат роботи нелінійного SVM-класифікатора

тивного класифікатора, представленого на персональному сайті В.Н. Вапніка [6].

Багатокласовий випадок зводять до задачі дихотомії: по черзі кожен клас порівнюють із сукупністю усіх інших або кожен клас із кожним.

Цей метод широко застосовують для розв'язання задач дистанційного зондування, оскільки за його допомогою фактично проводять картування пікселів на зображенні.

Втім оптимальність методу SVM не доведена. Також є проблематичним супроводження кожного результату класифікації оцінкою достовірності, що для статистичного підходу не є проблемою.

**3. Статистичний підхід (бейєсівський). Класичний підхід Бейєса.** В основі бейєсівського критерію лежать такі уявлення. Кожному класу  $Q_j$  відповідає певна щільність розподілу  $\xi$ :

$$p_j(\mathbf{x}) = p_j(x_1, x_2, \dots, x_m), \quad j = \overline{1, t}. \quad (5)$$

Класифікації підлягає об'єкт, що характеризується набором значень  $\mathbf{y} = \{y_1, y_2, \dots, y_m\}$  величини  $\xi$ . Задача полягає у визначенні класу, до якого належить  $\mathbf{y}$ .

Згідно з критерієм Бейєса [7, 8], класову належність  $\mathbf{y}$  визначають за максимальною апостеріорною ймовірністю

$$P\{\mathbf{y} \in Q_j\} = \frac{q_j p_j(\mathbf{y})}{\sum_{j=1}^t q_j p_j(\mathbf{y})} \quad (j = \overline{1, t}), \quad (6)$$

де  $p_j(\mathbf{x})$  – щільність  $m$ -вимірного розподілу  $\xi$  у класі  $Q_j$ ;  $q_j$  – апіорна ймовірність того, що  $\mathbf{y}$  належить  $Q_j$ .

Якщо апіорної інформації немає, найімовірніший клас визначаємо за максимальною ймовірністю (6) при  $q_j = 1/t$ :

$$P\{\mathbf{y} \in Q_j\} = \frac{p_j(\mathbf{y})}{\sum_{j=1}^t p_j(\mathbf{y})}, \quad (7)$$

тобто максимальною з величин  $p_j(\mathbf{y})$  ( $j = \overline{1, t}$ ). Отримали властивість оптимальності критерію Бейєса [9]: середньоочікувана ймовірність помилок є мінімальною.

Критерій (7) не враховує цін помилок, тобто втрат, пов'язаних із помилковим прийняттям або відхиленням гіпотез. Метод узагальнюємо, якщо ціни таких помилок є заданими. Критерій класифікації, за якого математичне очікування втрат є мінімальним, має такий вигляд: якщо

$$\sum_{i=1, i \neq k}^t q_i c_{ik} p_i(\mathbf{y}) < \sum_{i=1, i \neq j}^t q_i c_{ij} p_i(\mathbf{y}) \quad (j, k = \overline{1, t}, j \neq k), \quad (8)$$

то приймається гіпотеза  $\{\mathbf{y} \in Q_k\}$ . Тут  $c_{ij}$  – ціна помилкового віднесення  $\mathbf{y}$  до  $Q_j$ , тоді як в дійсності  $\mathbf{y}$  належить до  $Q_i$ .

Додаткова перевага бейєсівського підходу полягає у такому. Коли декілька класів,  $Q_{j_1}, Q_{j_2}, \dots, Q_{j_k}$ , є близькими за розподілами  $\xi$ , то достовірне віднесення до одного із групи таких однорідних класів неможливе. При цьому якщо об'єкт  $\mathbf{y}$  належить до такої групи, то кожна з імовірностей  $P\{\mathbf{y} \in Q_{j_i}\}$ ,  $i = \overline{1, k}$  може бути невеликою, але в сумі вони складають величину, настільки близьку до одиниці, що його практично достовірно можна віднести до групи класів. Критерій Бейєса дає змогу проводити класифікацію як за окремими класами, так і за їх об'єднаннями, причому на основі тих самих даних – імовірностей (7), підсумуванням останніх із зазначенням достовірності віднесення за таких об'єднань:

$$P\{\mathbf{y} \in Q_{j_1} \cup Q_{j_2} \cup \dots \cup Q_{j_k}\} = \sum_{i=1}^k P\{\mathbf{y} \in Q_{j_i}\}. \quad (9)$$

Отже, теоретично перевага бейєсівського правила безсумнівна. Проте є суттєва перешкода застосуванню описаного критерію в розглядуваній задачі, яка полягає в необхідності оцінювати вирази щільності  $p_j(\mathbf{x})$ . Це зазвичай не є простою задачею [10]. Як правило, можливо вказати тільки вибірки  $\tilde{Q}_j$  спостережень  $\xi$ , стосовно яких наперед відомо про їхню належність  $Q_j$ :  $\tilde{Q}_j = \{\mathbf{x}_1^{(j)}, \mathbf{x}_2^{(j)}, \dots, \mathbf{x}_{n_j}^{(j)}\}$ ,  $\mathbf{x}_i^{(j)} \in Q_j$ ,  $i = \overline{1, n_j}$ ,  $j = \overline{1, t}$  ( $n_j$  – кількість спостережень із  $Q_j$ ).

Спостереження  $\mathbf{y}_i^{(j)}$  слід використовувати як еталони, на яких і будемо критерій. Проблемою такого підходу є необхідність спиратися на часто сумнівну гіпотезу багатовимірного гаусівського закону розподілу і на додаток до нього – необхідність обернення коваріаційних матриць. Саме це змушує відмовлятися від теоретично оптимального критерію та звертатися до тих чи інших емпіричних алгоритмів.

Якщо параметричні сім'ї щільностей розподілів  $\xi$  в класах  $Q_j$  вважати заданими –

$$p_j(x_1, x_2, \dots, x_m; \alpha_1, \alpha_2, \dots, \alpha_{k_j}) = p_j(\mathbf{x}, \boldsymbol{\alpha}_j) \quad (j = \overline{1, t}), \quad (10)$$

то задача допускає розв'язок за параметричною класифікацією, що ґрунтується на використанні виразів щільностей (6). Позначивши через  $\{\tilde{\alpha}_1, \tilde{\alpha}_2, \dots, \tilde{\alpha}_{k_j}\}^T = \tilde{\boldsymbol{\alpha}}_j$  оцінки параметрів  $\{\alpha_1, \alpha_2, \dots, \alpha_{k_j}\}^T = \boldsymbol{\alpha}_j$ ,  $q_j$  – апіорні ймовірності належності  $\mathbf{y} = \{y_1, y_2, \dots, y_m\}$  до класів  $Q_1, Q_2, \dots, Q_t$  відповідно, отримаємо оцінки апостеріорних імовірностей, згідно із виразом (6):

$$\tilde{P}\{\mathbf{y} \in Q_j\} = \frac{q_j p_j(\mathbf{y}, \tilde{\boldsymbol{\alpha}}_j)}{\sum_{j=1}^t q_j p_j(\mathbf{y}, \tilde{\boldsymbol{\alpha}}_j)} \quad (j = \overline{1, t}). \quad (11)$$

Якщо апіорної інформації немає, то  $q_j$  беруть однаковими і такими, що дорівнюють  $1/t$ . На практиці зазвичай доводиться мати справу саме з такими умовами. Отже, при  $q_j = 1/t$  ( $j = \overline{1, t}$ )

$$P_j(\mathbf{y}) = \mathbf{P}\{\mathbf{y} \in Q_j\} \approx \frac{p_j(\mathbf{y}, \tilde{\mathbf{a}}_j)}{\sum_{j=1}^l p_j(\mathbf{y}, \tilde{\mathbf{a}}_j)}. \quad (12)$$

Такий розв'язок теоретично не виглядає складним, проте у реальній практиці виникають значні складнощі. Наприклад, у разі припущення нормального розподілу величини  $\xi$  щільності (10) мають вигляд [7]

$$p_j(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{m}{2}} |\mathbf{B}_j|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{m}_j)^T \mathbf{B}_j^{-1}(\mathbf{x} - \mathbf{m}_j)\right], \quad (13)$$

де  $\mathbf{m}_j$  – математичне сподівання величини  $\xi$  в  $Q_j$ ;  $\mathbf{m}_j = \{M_1^{(j)}, M_2^{(j)}, \dots, M_m^{(j)}\}$ ;  $\mathbf{B}_j$  – коваріаційна матриця,  $|\mathbf{B}_j|$  – її визначник. Оцінки математичних сподівань та коваріаційних матриць мають вигляд

$$\tilde{\mathbf{m}}_j = \{\tilde{M}_1^{(j)}, \tilde{M}_2^{(j)}, \dots, \tilde{M}_m^{(j)}\}, \quad \tilde{M}_k^{(j)} = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ik}^{(j)} \quad (k = \overline{1, m}); \quad (14)$$

$$\tilde{\mathbf{B}}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} (\mathbf{x}_i^{(j)} - \tilde{\mathbf{m}}_j)(\mathbf{x}_i^{(j)} - \tilde{\mathbf{m}}_j)^T. \quad (15)$$

Підставивши ці оцінки у вираз (13), отримаємо оцінки щільності розподілу  $p_j(\mathbf{x})$ .

Згідно з рівностями (13)–(15) оцінки апостеріорних імовірностей мають вигляд:

$$\tilde{P}_j(\mathbf{y}) = \frac{|\tilde{\mathbf{B}}_j|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \tilde{\mathbf{m}}_j)^T \tilde{\mathbf{B}}_j^{-1}(\mathbf{y} - \tilde{\mathbf{m}}_j)\right]}{\sum_{i=1}^l |\tilde{\mathbf{B}}_i|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \tilde{\mathbf{m}}_i)^T \tilde{\mathbf{B}}_i^{-1}(\mathbf{y} - \tilde{\mathbf{m}}_i)\right]} \quad (j = \overline{1, l}). \quad (16)$$

Головна проблема – потреба мати значні обсяги еталонних вибірок, особливо за великої розмірності через необхідність отримання стійких оцінок обернених матриць  $\tilde{\mathbf{B}}_j^{-1}$ . Тому доводиться робити не завжди обґрунтовані припущення про такі особливості багатовимірних розподілів, які дають змогу полегшити розв'язок і поліпшити його стійкість. Зокрема, таке припущення може полягати в тому, що компоненти  $\xi_1, \xi_2, \dots, \xi_m$  є взаємно незалежними в усіх класах. Однак таке припущення обмежує використання корельованих компонент, зокрема, близьких між собою каналів. За гіпотези некорельованості щільності сумісних розподілів  $\xi_1, \xi_2, \dots, \xi_m$  у кожному класі  $Q_j$  являють собою добутки одновимірних щільностей  $\xi_i$  ( $i = \overline{1, m}$ ):

$$p_j(\mathbf{x}) = p_j(x_1, x_2, \dots, x_m) = p_1^{(j)}(x_1) p_2^{(j)}(x_2) \dots p_m^{(j)}(x_m), \quad (17)$$

де  $p_i^{(j)}(x_i)$  – щільність розподілу  $\xi_i$  в  $Q_j$ .

За гіпотези нормального розподілу  $\xi_i$

$$p_i^{(j)}(x_i) = \frac{1}{\sqrt{2\pi\sigma_i^{(j)}}} \exp\left[-\frac{1}{2} \frac{(x_i - M_i^{(j)})^2}{(\sigma_i^{(j)})^2}\right], \quad (18)$$

де  $M_i^{(j)}$  і  $\sigma_i^{(j)}$  – математичне сподівання та середньоквадратичний відхил  $\xi_i$  в  $Q_j$ . Обчисливши їх

оцінки за спостереженнями  $x_{it}^{(j)}$  з вибірок  $\tilde{Q}_j - \tilde{M}_i^j$  та  $\tilde{\sigma}_i^{(j)} = \sqrt{\tilde{b}_i^{(j)}}$  згідно з (15), отримаємо можливість оцінити одновимірні щільності (18). Таким чином, у випадку незалежності величин  $\xi_i$  оцінками апостеріорних ймовірностей (12) будуть

$$\tilde{\mathbf{P}}\{\mathbf{y} \in Q_j\} = \tilde{P}_i(\mathbf{y}) = \frac{\prod_{i=1}^m \tilde{p}_i^{(j)}(y_i)}{\sum_{i=1}^l \prod_{i=1}^m \tilde{p}_i^{(i)}(y_i)} \quad (j = \overline{1, l}), \quad (19)$$

де  $\tilde{p}_i^{(j)}$  визначають з (18) за умов  $M_i^{(j)} = \tilde{M}_i^{(j)}$  та  $\sigma_i^{(j)} = \tilde{\sigma}_i^{(j)}$ .

Інше можливе припущення полягає у тому, що величини  $\xi_i$  є залежними, але коваріаційні матриці однакові:  $\mathbf{B}_1 = \mathbf{B}_2 = \dots = \mathbf{B}_m = \mathbf{B}$ . У цьому випадку замість оцінок матриць  $\tilde{\mathbf{B}}_j$ , які застосовані у виразі (13), можна використовувати одну оцінку  $\tilde{\mathbf{B}}$ , обчисливши її як середньозважену із  $\tilde{\mathbf{B}}_j$ :

$$\tilde{\mathbf{B}} = \frac{1}{N} \sum_{j=1}^l n_j \tilde{\mathbf{B}}_j \quad \left(N = \sum_{j=1}^l n_j\right). \quad (20)$$

Тоді оцінки ймовірностей (16) мають вигляд

$$P_j(\mathbf{y}) = \frac{\exp\left[-\frac{1}{2}(\mathbf{y} - \tilde{\mathbf{m}}_j)^T \tilde{\mathbf{B}}^{-1}(\mathbf{y} - \tilde{\mathbf{m}}_j)\right]}{\sum_{i=1}^l \exp\left[-\frac{1}{2}(\mathbf{y} - \tilde{\mathbf{m}}_i)^T \tilde{\mathbf{B}}^{-1}(\mathbf{y} - \tilde{\mathbf{m}}_i)\right]} \quad (j = \overline{1, l}). \quad (21)$$

**Непараметрична класифікація.** Автори дотримуються точки зору, що вищеописані проблеми не стають задавою для методу непараметричної класифікації [11]. У нього немає обмежень, які пов'язані із законом розподілу або корельованістю. Разом з тим метод побудований за принципом оптимального бейєсівського критерію і тому має потенціал перелічених вище переваг.

Згідно з цим методом, імовірність належності до класу  $Q_j$  за відсутності апріорних імовірностей оцінено у вигляді

$$P_j(\mathbf{y}) = \mathbf{P}\{\mathbf{y} \in Q_j\} = \frac{P_j(\Omega)}{\sum_{j=1}^l P_j(\Omega)} \quad (j = \overline{1, l}), \quad (22)$$

де  $P_j(\Omega) = \mathbf{P}\{\xi \in \Omega / \xi \in Q_j\}$  – імовірність того, що значення величин  $\xi_1, \xi_2, \dots, \xi_m$  під час спостережень їх у класі  $Q_j$  утворюють точку  $m$ -вимірного простору, яка потрапляє в окіл  $\Omega$  навколо  $\mathbf{y}$ . Оцінками величин  $P_j(\Omega)$  ( $j = \overline{1, l}$ ) є

$$\tilde{P}_j(\Omega) = \frac{n_j(\Omega)}{n_j} \quad (j = \overline{1, l}), \quad (23)$$

де  $n_j(\Omega)$  – кількість  $m$ -вимірних еталонних спостережень  $\mathbf{x}_i^{(j)}$  із  $Q_j$ , які потрапили в  $\Omega$ . Найпростіша реалізація правила, що описується, полягає у визначенні груп спостережень  $\mathbf{x}_i^{(j)}$  у кожній вибірці  $\tilde{Q}_j$ , найближчих за відстанню між  $\mathbf{y}$  та

$x_i^{(j)}$  задалегідь сформованої структури. Такою структурою може бути

$$\rho(x_i^{(j)}, y) = \sqrt{\sum_{l=1}^m \alpha_l^2 (x_{il}^{(j)} - y_l)^2}, \quad (24)$$

де  $\alpha_l$  – коефіцієнти масштабу, які вибирають відповідно до варіації кожної компоненти у класах  $Q_i$ . Зокрема, прийнятні результати може забезпечити вибір  $\alpha_l$ , обернено пропорційних середньому стандартному відхиленню  $l$ -ї компоненти:

$$\alpha_l = \frac{1}{\bar{\sigma}_l} \left( \sum_{l=1}^m \frac{1}{\bar{\sigma}_l} \right)^{-1}. \quad (25)$$

Можна обговорювати й інші метрики, наприклад, середньозважений абсолютний відхил

$$\rho_a(x_i^{(j)}, y) = \sum_{l=1}^m \alpha_l |x_{il}^{(j)} - y_l|$$

або зважене максимальне розходження

$$\rho_a(x_i^{(j)}, y) = \max_{l=1, m} \alpha_l |x_{il}^{(j)} - y_l|.$$

Задавши фіксований рівень  $\beta < 1$ , обчислимо  $f_j$  – округлені до цілого значення  $\beta \cdot n_j$ . Далі у кожній вибірці  $Q_j$  визначаємо  $f_j$  найближчих до  $y$  за відстанню (24) векторів  $x_i^{(j)}$ , а серед них найдалший  $x_d^{(j)}$ . Позначивши  $\rho^* = \min \rho(x_d^{(j)}, y)$ , знаходимо у кожній вибірці  $Q_j$  кількість спостережень  $n_j^*$ , якщо  $\rho(x_i^{(j)}, y) < \rho^*$ . Імовірності (22) оцінюємо у вигляді

$$\tilde{P}_j(y) = \frac{n_j^*}{n_j} \left( \sum_{j=1}^t \frac{n_j^*}{n_j} \right)^{-1}, \quad (26)$$

а клас, до якого відносять  $y$ , визначаємо за найвищою оцінкою. Групу найімовірніших класів виявляємо за найвищими значеннями оцінок (26), сума яких оцінює ймовірність того, що  $y$  належить до цієї групи.

Для отримання прозорості в імовірнісному розумінні характеристики ефективності критерію використовують результати класифікації еталонних об'єктів із кожного класу. Класифікувавши кожен такий об'єкт, обчислюємо оцінки ймовірності помилки  $\tilde{q}_{ij} = n_{ij} / n_i$ , де  $n_i$  – кількість еталонних об'єктів із  $Q_i$ ;  $n_{ij}$  – ті з них, що були віднесені до класу  $Q_j$ . До того ж ці оцінки дають змогу виділити найближчі класи, які можна розділити із малою достовірністю.

Емпіричний критерій класифікації може бути так улаштований, що еталонні об'єкти можна впізнавати безпомилково за їхньою специфічністю, а не за належністю до класу. Така ситуація, наприклад, може виникнути у разі використання методу найближчого сусіда. Для того щоб позбавитися від цього цілком прогнозованого ефекту хибної оптимальності, можна рекомендувати спосіб поточного виключення з ета-

лонних вибірок того об'єкта, який підлягає тестовій класифікації.

Для зазначеного критерію не створює проблем надмірна вимірність та не існує ефекту перенавчання, як в інших методах класифікації.

Слід згадати про ще одну можливість підвищення ефективності вирішальних правил – послідовної (ієрархічної) класифікації. Класифікацію виконують не одразу на класи  $Q_1, Q_2, \dots, Q_t$ , а на певні їх об'єднання  $Q_1^0, Q_2^0, \dots, Q_k^0$  ( $k < t$ ), що складаються з одного або кількох класів  $Q_i$ . При цьому об'єднують найближчі класи так, щоб отримати відокремлені групи класів. Для класифікації на  $Q_1^0, Q_2^0, \dots, Q_k^0$  вибирають систему показників. Після цього етапу класифікації, якщо вектор віднесено до об'єднання  $Q_j^0$ , визначають його належність до одного з дрібніших об'єднань або безпосередньо класів  $Q_i$ , які складають  $Q_j^0$ .

Для цього знову підбирають інформативні показники і систему підрахунку їх інформативності. Такий підхід дає змогу підвищити ефективність класифікації у певній послідовності, наприклад “підтоплення критичні – некритичні”, потім кожен клас поділяють на дрібніші.

Найскладнішою ситуацією для усіх без винятку критеріїв класифікації є задача з класами, які займають проміжне положення між іншими. Ілюстрацією є рис. 3 – клас  $Q_2$  займає проміжне положення між  $Q_1$  та  $Q_3$ , що практично унеможливає прийняття рішення про належність об'єкта класу  $Q_2$  на всій області розв'язків, оскільки там  $p_2(x) < \max(p_1(x), p_3(x))$ .

Використання методу непараметричної класифікації внаслідок близькості до оптимального критерію дає змогу певною мірою сподіватися на позбавлення від цього неприємного ефекту.

**Висновки.** Методи розв'язання задачі класифікації пікселів багатоканальних космічних знімків за станом підтоплення можна поділити на емпіричні та параметричні. Метод лінійних роздільних площин не можна вважати оптимальним у розумінні мінімуму середньоочікуваних помилок класифікації. Водночас теоретично оптимальний бейєсівський класифікатор обмежений у своїй параметричній реалізації. Запропоновано непараметричний підхід до побудови бейєсівського критерію класифікації, показано переваги порівняно з емпіричним підходом.

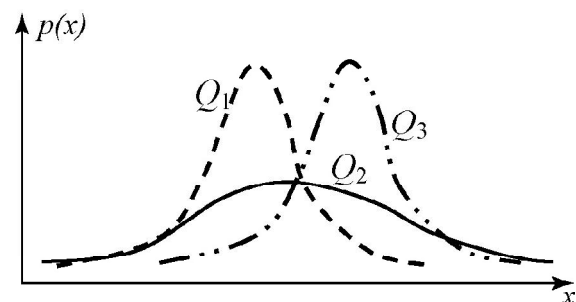


Рис. 3. Приклад проміжного розподілу показника в одному з класів ( $Q_2$ )

1. Проект “Державні будівельні норми “Інженерний захист від підтоплення ґрунтовими водами територій та споруд”. – <http://uwodgeo.org.ua/normdoc/DBN%20V11%202007.zip>
2. Крынов Е.Л. Спектральная отражательная способность природных образований. – М.: Изд-во АН СССР, 1947. – 272 с.
3. Гонсалес Р., Вудс Р. Цифровая обработка изображений / Пер. с англ.; Под ред. П.А. Чочиа. – М.: Техносфера, 2005. – 1072 с.
4. Vapnik V.N. The Nature of Statistical Learning Theory (Statistics for Engineering and Information Science). – 2nd Edition – New York: Springer Verlag, 2000. – 314 p.
5. Nahn C., Gloaguen R. Estimation of soil types by non linear analysis of remote sensing data // Nonlinear Processes in Geophysics. – 2008. – 15. – P. 115–126.
6. Computer learning research centre. – <http://svm.cs.rhul.ac.uk/pagesnew/GPat.shtml>
7. Андерсон Т. Введение в многомерный статистический анализ. – М.: Физматгиз, 1963.
8. Жуков М.Н. Статистичний аналіз геологічних даних. – К.: Віпол, 1995. – 551 с.
9. Жуков Н.Н. Вероятностно-статистические методы анализа геолого-геофизической информации. – К.: Вища шк., 1975. – 303 с.
10. Куликович А.Е., Якимчук Н.А. Проблемы геоинформатики. Ч. 1. – Киев: Центр менеджмента и маркетинга в области наук о Земле, 2002. – 77 с.
11. Никитин А.А. Статистические методы выделения геофизических аномалий. – М.: Недра, 1979. – 280 с.
12. Крамер Г. Математические методы статистики. – М.: Мир, 1975. – 648 с.

Надійшла до редакції 17.02.2009 р.

*М.Н. Жуков, А.М. Макаренко*

### БАГАТОВИМІРНА КЛАСИФІКАЦІЯ В ЗАДАЧАХ МОНІТОРИНГУ ПІДТОПЛЕНЬ ЗАСОБАМИ ДИСТАНЦІЙНОГО ЗОНДУВАННЯ ЗЕМЛІ

Подано оцінку можливостей бейєсівського підходу порівняно з методом опорних векторів (SVM) для вирішення задачі класифікації земель за станом підтоплення. Обговорено способи реалізації теоретично оптимального критерію у розумінні мінімуму математичного очікування втрат або помилок класифікації. Зроблено висновок на користь методу непараметричної класифікації, алгоритм якої описано. Розглянуто способи підвищення достовірності класифікації.

*Н.Н. Жуков, А.Н. Макаренко*

### МНОГОМЕРНАЯ КЛАССИФИКАЦИЯ В ЗАДАЧАХ МОНИТОРИНГА ПОДТОПЛЕНИЙ СРЕДСТВАМИ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ

Дается оценка возможностей бейесовского подхода в сравнении с методом опорных векторов (SVM) для решения задачи классификации земель по состоянию подтопления. Обсуждаются пути реализации теоретически оптимального критерия в понимании минимума математического ожидания потерь или ошибок классификации. Сделан вывод в пользу метода непараметрической классификации, алгоритм которой описан. Рассматриваются приемы улучшения достоверности классификации.