

УДК 528.854.4:574.9

ОСОБЛИВОСТІ РОЗРАХУНКУ ІНДЕКСУ ВИДОВОГО РІЗНОМАНІТТЯ ЗА РЕЗУЛЬТАТАМИ СТАТИСТИЧНОЇ КЛАСИФІКАЦІЇ АЕРОКОСМІЧНИХ ЗНІМКІВ

Станкевич С.А., Козлова А.О.

Аналіз і подальша кількісна оцінка різноманіття біосистем різних рівнів інтеграції передбачають застосування математично виражених показників. У дослідженнях видового різноманіття, зокрема при його оцінці, широко використовується ентропійний індекс Шеннона [1]:

$$I = - \sum_{j=1}^m p_j \log_2 p_j, \quad (1)$$

де m - кількість видів, p_j - значущість виду j .

Здатність індексу враховувати не лише багатство видів, а й рівномірність розподілу цих видів за будь-якою своєю ознакою, обумовлює трудомісткість його розрахунку.

Залучення даних дистанційного зондування та сучасних цифрових технологій обробки аерокосмічних зображень, на нашу думку, може додати ряд переваг у процес розрахунку індексу видового різноманіття за Шенноном: значно розширити територію дослідження, відобразити реальну інформацію щодо просторового розподілу рослинних угруповань та надавати можливість її постійного оновлення, а представлення даних у цифровому вигляді забезпечує їх ефективну обробку. При цьому слід пам'ятати, що точність кінцевого результату обумовлена якістю інтерпретації аерокосмічного зображення, на яку значно впливають особливості обраного алгоритму класифікації.

Більшість відомих алгоритмів класифікації багатоспектральних цифрових аерокосмічних зображень засновано на статистичних моделях, оскільки земні утворення, що підлягають дистанційному аерокосмічному спостереженню, мають широке різноманіття власних та наведених спектральних характеристик, характер яких в загальному випадку невідомий або змінюється за невідомими факторами [2]. Принципова стохастичність корисних багатоспектральних оптичних сигналів земних утворень обумовлюється такими факторами, як варіації структурного та геохімічного складу об'єктів спостереження, зміни зовнішніх геофізичних полів, погодних умов, зміни оптичних властивостей середовища та освітлення, похибки калібрування та власні шуми каналів реєстрації, тощо.

Подальші перетворення оптичних сигналів при проходженні крізь атмосферу, оптичні системи, фотоприймальні пристрої, відеотракти, аналогово-цифрові та цифро-аналогові перетворювачі апаратури дистанційного спостереження вносять до сигналів додаткові зміни, характеристики яких далеко не завжди відомі або можуть бути враховані. Тому натівним математичним апаратом опису оптичних сигналів

земних утворень здається теорія ймовірності та прикладна математична статистика [3].

Все різноманіття статистичних методів класифікації багатоспектральних цифрових аерокосмічних зображень в першому наближенні можливо звести до класичних, пов'язаних з оцінюванням параметрів імовірнісних розподілів за статистичними вибірками [4]; непараметричних, які треба застосовувати в разі, коли реальні дані не підпорядковуються будь-яким параметричним сімействам [5]; нечітко-статистичних, зокрема інтервальних, які більш стійки до похибок та потребують менших обчислювальних витрат [6]; нечислових або експертних, які безпосередньо оброблюються механізмами логічного виводу або зводяться до числових процедурами багатовимірної шкалування [7].

При застосуванні класичних статистичних методів використовуються:

- статистичні оцінки імовірнісних характеристик випадкових процесів, що описують гіперспектральні оптичні сигнали та дозволяють побудувати параметричні оцінки густини імовірності кожного з них;

- алгоритми багатоальтернативної перевірки гіпотез про належність зареєстрованого оптичного сигналу тій чи іншій випадковій послідовності, та, відповідно, тому чи іншому класу об'єктів дистанційного спостереження.

Нехай на s -вимірному просторі X багатоспектральних відліків $X = \{x_\lambda\}$, $\lambda = 1 \dots s$ завдано m класів оптичних сигналів $X^{(j)} = \{x_\lambda^{(j)}\}$, $j = 1 \dots m$. Задача полягає в тому, щоб за результатами аналізу конкретної реалізації випадкового процесу прийняти гіпотезу про належність гіперспектрального оптичного сигналу одному з цих класів. Звичайно простір X розбивається на m областей $X^{(m)}$, які не перекриваються.

Якщо відомий аналітичний опис густини імовірності розподілу $f(X, \Theta)$ класів $X^{(j)}$, $j = 1 \dots m$ в просторі X з вектором параметрів Θ та завдані апріорні імовірності гіпотез $p_j = P[X \in X^{(j)}]$, $j = 1 \dots m$, то за формулою Байеса можна знайти апостеріорні імовірності гіпотез належності довільної гіперспектральної вибірки X k -му класу $X^{(k)}$:

$$P(X \in X^{(k)}) = \frac{p_k f(X, \Theta^{(k)})}{\sum_{j=1}^m p_j f(X, \Theta^{(j)})}, \quad j = 1 \dots m. \quad (2)$$

Із (1) витікає наступний оптимальний за критерієм максимуму апостеріорної імовірності алгоритм прийняття рішення про належність X класу $X^{(k)}$ [8]:

$$p_k f(X, \Theta^{(k)}) = \max_j p_j f(X, \Theta^{(j)}) . \quad (3)$$

В багатьох випадках правомірним є опис густини імовірності класів багатовимірними Гаусівськими законами розподілу [9]:

$$\Theta^{(j)} = \{X^{(j)}, V^{(j)}\},$$

$$f(X, \Theta^{(j)}) = \frac{1}{\sqrt{(2\pi)^n \det V^{(j)}}} \exp \left[-\frac{(X - X^{(j)})^T (V^{(j)})^{-1} (X - X^{(j)})}{2} \right], \quad (4)$$

де $X^{(j)} = \{x_\lambda^{(j)}\}$, $\lambda = 1 \dots s$ – вектор середніх значень j -го класу, $V^{(j)} = \{c_{ll}^{(j)}\}$, $\lambda, l = 1 \dots s$ – коваріаційна матриця j -го класу.

Непараметричні статистичні методи використовуються, коли аналітичне уявлення розподілу густини імовірності класів оптичних сигналів невідомо або взагалі неможливе внаслідок специфіки об'єктів спостереження, принципової нелінійності каналів реєстрації гіперспектральної інформації, значних похибок, викривлень та викидань даних, шумів, тощо.

Найбільш часто використовуються знакові, порядкові та рангові непараметричні статистичні методи [10]. В знакових методах розглядаються знакові вектори $\text{sign}(X-X^{(j)})$ гіперспектральних відліків

$$\text{sign}(X-X^{(j)}) = \{\text{sign}(x_\lambda - x_\lambda^{(j)})\}, \quad \lambda = 1 \dots s, \quad (5)$$

де

$$\text{sign } x = \begin{cases} 1 & \text{при } x > 0 \\ 0 & \text{при } x = 0 \\ -1 & \text{при } x < 0 \end{cases} \quad (6)$$

знакова функція. Якщо вектори X та $X^{(j)}$ є різними реалізаціями одного випадкового процесу, то функція розподілу густини імовірності знакового вектора (5) буде симетричною відносно нуля $f(x) = f(-x)$ та імовірність випадкового збігу знаків багатоспектральних векторів складе $\left(\frac{1}{2}\right)^s$.

Знаковий алгоритм прийняття рішення про належність X класу $X^{(k)}$ прийме вигляд

$$\sum_{\lambda=1}^s \text{sign}(x_\lambda - x_\lambda^{(k)}) = \min_j \sum_{\lambda=1}^s \text{sign}(x_\lambda - x_\lambda^{(j)}). \quad (7)$$

Треба відзначити, що алгоритм (7) неінваріантний до постійного зсуву значень елементів вектора X , та потребує радіометричної корекції гіперспектральних зображень.

В порядкових методах або розглядаються ті чи інші непараметричні кореляції, наприклад, коефіцієнт Кендалла τ , або критеріальні показники статистичної значимості розбіжностей між векторами гіперспектральних відліків [11]. Всі вони базуються на будівні варіаційного ряду – перегрупованої в порядку зростання значень елементів вибірки $X \rightarrow Z$, $Z = \{z_j\}$, $z_j \leq z_k$ при $j < k$. Коефіцієнт кореляції Кендалла визначається за формулою

$$\tau(X, X^{(j)}) = \frac{4 \sum_{l=1}^s G[z_l(x_i), z_l^{(j)}]}{s(s-1)} - 1, \quad (8)$$

де $G[z_l, z_l^{(j)}]$ – кількість більших, ніж $z_l^{(j)}$, елементів послідовності Z з порядковими номерами більшими, ніж l . Порядковий алгоритм прийняття рішення про належність X класу $X^{(k)}$ за коефіцієнтом Кендалла буде

$$\tau(X, X^{(k)}) = \max_j \tau(X, X^{(j)}) \quad (9)$$

В рангових методах виконується перетворення вхідних вибірок гіперспектральних відліків до їх рангових статистик $\text{rank } X^{(j)} = \{\text{rank } x_\lambda^{(j)}\}$, де $\text{rank } x_\lambda^{(j)}$ – ранг елемента $x_\lambda^{(j)}$, тобто його порядковий номер у варіаційному ряді $Z^{(j)} = \{z_i^{(j)}\}$. Рангові непараметричні статистики оперують рангами випадкових величин кожної вибірки. Ранговий критерій Вілсона [11] використовує тестові статистики вигляду

$$\omega(X, X^{(j)}) = \frac{\sum_{\lambda=1}^s \text{rank } x_\lambda - x_\omega}{\sqrt{\sigma_\omega^2}} \quad (10)$$

де

$$x_\omega = \frac{s \cdot (2s + 1)}{2} \quad (11)$$

рангове середнє варіаційного ряду, та

$$\sigma_\omega^2 = \frac{s^2(2s + 1)}{12} \quad (12)$$

дисперсія варіаційного ряду. Ранговий алгоритм прийняття рішення про належність X класу $X^{(k)}$ за критерієм Вілсона може бути записаний як

$$\omega(X, X^{(k)}) = \min_j \omega(X, X^{(j)}) \quad (13)$$

Нечітко-статистичні алгоритми класифікації багатоспектральних оптичних сигналів використовують їх функції належності $\mu(x_\lambda)$, $\lambda = 1 \dots s$ [12]. Відповідний алгоритм прийняття рішення про належність багатоспектрального оптичного сигналу X класу $X^{(k)}$ запишеться як

$$\mu(X \leftrightarrow X^{(k)}) = \max_j \&_{\lambda=1}^s \mu(x_\lambda) \leftrightarrow \mu(x_\lambda^{(j)}) \quad (14)$$

де \leftrightarrow – символ нечіткої еквівалентності, $\&$ – оператор нечіткої кон'юнкції.

Сучасні цифрові технології обробки цифрових аерокосмічних зображень надають значні можливості статистичного аналізу багато- та гіперспектральних даних з урахуванням їх детальних статистичних властивостей [13].

Особливостями всіх розглянутих статистичних алгоритмів класифікації є наявність деякої кількісної оцінки належності поточного пікселя цифрового аерокосмічного зображення кожному класу земних утворень. Для класичних

статистичних алгоритмів – це імовірність (2) або її густина (4), для непараметричних статистичних алгоритмів – це знакова розбіжність (7), коефіцієнт Кендалла (8) або ранговий критерій Вілсона (10), для нечітко-статистичних алгоритмів – це нечітка кон'юнкція (14), тощо. Причому, при застосуванні будь-якого прикінцевого вирішального правила (3), (7), (9), (13) або (14) про остаточне віднесення пікселя до класу k ця інформація спростовується [14]. Між тим, її використання при наступній обробці цифрових аерокосмічних зображень, що засновується на результатах класифікації, могло б надати додаткові дані та підвищити загальну точність.

Розглянемо аерокосмічне зображення з n пікселів, яке містить m класів, які враховуються при оцінюванні різноманіття.

Оцінка кількості пікселів n_k в класі k може бути розрахована за формулою:

$$n_k = \sum_{i=1}^n p_i^{(k)}, \quad (15)$$

де $p_i^{(k)}$ – апостеріорна імовірність класу k в пікселі i . Її можна оцінити за баєсівським підходом (2) як

$$p_i^{(k)} = \frac{f_i^{(k)}}{\sum_{j=1}^m f_i^{(j)}} \quad (16)$$

де $f_i^{(j)}$ – функція належності пікселя i класу j , одержана за результатами роботи алгоритму класифікації.

Оцінка загальної кількості пікселів по всіх класах n здійснюється за формулою:

$$n = \sum_{j=1}^m n_j, \quad (17)$$

Таким чином, вказані особливості статистичних алгоритмів можуть бути враховані при оцінюванні шенонівського індексу видового різноманіття шляхом обчислення за формулою:

$$I = \sum_{k=1}^m \frac{n_k}{n} \cdot \log_2 \frac{n}{n_k}, \quad (18)$$

Для практичного випробування описаного вище підходу, спочатку за методом максимальної вірогідності було проведено класифікацію фрагмента космічного знімку Landsat ETM+ (05.06.2000) на територію Дунаєвського району Хмельницької області. Класи призначалися за видами судинних рослин, які формують верхній ярус у фітоценозах досліджуваної території (рис.1).

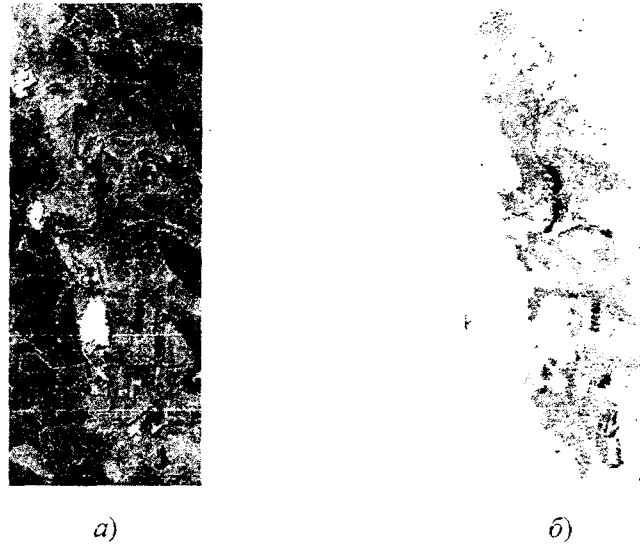


Рис. 1. Фрагмент багатозонального космічного знімку Landsat ETM+ (05.06.2000) на територію Дунаєвського району Хмельницької області (а) та одержаний на його основі класифікаційний тематичний растровий шар видового складу вищих судинних рослин (б)



Умовні позначення: 1 – сосна, 2 – ялина, 3 – дуб, 4 – береза, 5 – граб, 6 – території, не вкриті лісом, 7 – вода

Одержаний таким чином тематичний растровий шар являє собою підґрунтя для подальшої кількісної оцінки видового різноманіття судинних рослин даної території за ентропійним індексом Шеннона. Під час розрахунку індексу, в якості значущості виду було вказано частку площі, яку займає вид на досліджуваній території (табл.1). Згідно (1) індекс Шеннона для видового різноманіття вищих судинних рослин в межах фрагменту досліджуваної території становить 1,4356095.

Врахування даних про розподіл ймовірностей дозволяє підвищити точність обчислення значущості кожного з розглянутих класів (табл.1). В такому разі, індекс Шеннона обчислений за формулою (18) становитиме 1,5532077.

Таблиця 1.

Значущості видів досліджуваної території, за результатами класифікації на основі методу максимальної вірогідності (а) та з урахуванням особливостей розподілу ймовірностей (б)

| № | Назва виду | Значущість а) | Значущість б) | Відносне відхилення |
|----|-----------------------------------|------------------|------------------|------------------------|
| 1. | Сосна (<i>Pinus sylvestris</i>) | 0,0340394 | 0,0334447 | - 0,0177816 |
| 2 | Ялина (<i>Picea abies</i>) | 0,0358128 | 0,0421514 | 0,1503770 |
| 3. | Дуб (<i>Quercus robur</i>) | 0,5541872 | 0,5232288 | - 0,0591680 |
| 4. | Береза (<i>Betula pendula</i>) | 0,0159113 | 0,0373204 | 0,5736568 |
| 5. | Граб (<i>Carpinus betulus</i>) | 0,3600493 | 0,3638548 | 0,0104588 |

Отже, беручи до уваги особливості розрахунку індексу Шеннона на основі статистичної класифікації аерокосмічних знімків, можна підвищити точність кінцевого результату на 5–15 %. Викладений підхід є особливо актуальним при використанні аерокосмічних знімків невисокого просторового розрізнення.

Безперечно, можливість розрахунку індексу Шеннона (чи будь-якого іншого кількісного показника видового різноманіття) на основі матеріалів аерокосмічної зйомки мають певні обмеження. Проте, їх залучення є цілком слушним для дослідження й оцінки видового різноманіття рослинних угруповань, види яких добре інтерпретуються на аерокосмічних зображеннях. Представлений підхід може бути особливо корисним при складанні відповідних тематичних карт.

Список літератури

1. Протасов А. А. Биоразнообразие и его оценка: Концептуальная диверсикология.- К.: НАН Украины; Институт гидробиологии. 2002.- 107 с.
2. Alföldi T.T. Introduction to Digital Images and Digital Analysis Techniques: A Basic Course for the Appreciation of Digital Analysis of Remotely Sensed Multispectral Data.- Ottawa: Canada Centre for Remote Sensing, 1996.- 46 p.
3. Сотский Б.А. О статистической классификации оптических полей // Оптика и спектроскопия, 1994.- Т.76.- № 6.- С.986-987.
4. Патрик Э. Основы теории распознавания образов: Пер. с англ.- М.: Сов. Радио, 1980.- 408 с.
5. Холландер М., Вулф Д. Непараметрические методы статистики: Пер. с англ.- М.: Финансы и статистика, 1985.- 208 с.
6. Станкевич С.А. Нечетко-интервальный алгоритм распознавания // Информационные технологии получения и обработки данных воздушной разведки.- Киев: МО Украины, 1994.- С.24-26.
7. Саати Т., Кернс К. Аналитическое планирование. Организация систем: Пер. с англ.- М.: Радио и связь, 1991.- 224 с.
8. Левин Б.Р. Теоретические основы статистической радиотехники.- М.: Радио и связь, 1989.- 656 с.
9. Gomes A., Marcal A.R. Land Cover Revision Through Object Based Supervised Classification of ASTER Data // Proceedings of the ASPRS 2003 Annual Conference.- Anchorage, 2003.- P.45-49.
10. Тартаковский А.Г. Последовательные методы в теории информационных систем.- М.: Радио и связь, 1991.- 282 с.
11. Рунион Р. Справочник по непараметрической статистике.- М.: Финансы и статистика, 1982.- 562 с.
12. Haykin S., Kosko B. Intelligent Signal Processing.- San Diego: IEEE Press, 2001.- 576 pp.
13. Moreno J.F., Gonzalez M.-C., Alonso L., Hegarat-Masclé S.L., Alem F., Normand M. Survey of Remote Sensing Data Analysis Methods / Report for European Research Project AIMWATER by Contract ENV4-CT98-0740.- Valencia: CNRS/CETP, 1999.- 52 p.
14. Попов М.А., Лялько В.И., Подорван В.Н., Сахацкий А.И. Методика классификации природных объектов на многоспектральных космических изображениях на основе последовательного слияния информации // Материалы Второй открытой Всероссийской научной конференции "Современные проблемы дистанционного зондирования Земли из космоса".- М.: ИКИ РАН, 2004.- С.46-47.

Статья поступила в редакцию 28.04.06