

# КОНТРОЛИРУЕМАЯ КЛАССИФИКАЦИЯ ГИПЕРСПЕКТРАЛЬНЫХ ДАННЫХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ

А. А. Бучнев, В. П. Пяткин

*Институт вычислительной математики и математической геофизики СО РАН, 630090, Новосибирск*

УДК 528.852

Рассматривается система классификации с обучением гиперспектральных данных дистанционного зондирования Земли (ДЗЗ). Оценивается применимость классификатора Байеса для нормально распределенных векторов признаков к обработке этих данных. Описана реализация классификатора спектрально-углового картирования.

**Ключевые слова:** гиперспектральные данные дистанционного зондирования, классификация с обучением, классификатор Байеса, классификатор спектрально-углового картирования.

## Введение

Известно (см., например, Шовенгердт [1]), что непосредственное использование традиционных алгоритмов для обработки данных гиперспектральных сканеров приводит к определенным трудностям, связанным с большой размерностью векторов измерений. Ранее в ИВМиМГ СО РАН совместно с НИЦ «Планета» в рамках создания программного комплекса PlanetMonitoring по обработке данных дистанционного зондирования Земли (ДЗЗ) была разработана система классификации с обучением (контролируемой классификации) [2]. Эта система классификации реализует стратегию максимального правдоподобия Байеса для нормально распределенных векторов признаков в классах. Оценки параметров классификатора (векторов средних и ковариационных матриц в классах) получаются на основе обучающих выборок.

Этот классификатор, разработанный для обработки данных ДЗЗ с относительно небольшим количеством спектральных диапазонов (мультиспектральные данные), был достаточно легко адаптирован для обработки гиперспектральных образов, полученных сканерами с высоким спектральным разрешением. Однако такое прямое использование классификатора сопровождается следующими негативными обстоятельствами:

1. Высокие вычислительные затраты, обусловленные большой размерностью векторов признаков  $N$  (например, сканер ГСА спутника Ресурс-П № 3 имеет  $N = 255$  спектральных диапазонов). Расстояние Махалонобиса, используемое для оценки близости векторов признаков к центрам классов, является значением квадратичной формы с весовой матрицей размерности  $N \times N$ .

2. Требуются большие объемы обучающих данных. Для того, чтобы избежать сингулярностей при вычислении матриц, обратных к ковариационным, минимальный объем обучающих выборок в классах должен равняться  $N + 1$  вектору. Известно, что для получения хороших оценок ковариационных матриц в мультиспектральных данных ( $N \approx 10$ ) объем обучающей выборки  $V$  в классе должен быть таким, чтобы выполнялось следующее условие для соотношения размерности [3]:  $V/N \geq 100$ . Ясно, что для выполнения подобного условия в гиперспектральных данных ( $N \approx 100$ ) объем обучающей выборки должен быть увеличен не менее, чем на порядок.

Для преодоления этих затруднений необходимо иметь возможность извлечения из гиперспектральных данных признаков, позволяющих построить набор векторов, размерность которых соответствует размерности мультиспектрального образа.

С этой целью полный набор спектральных диапазонов разбивается на блоки (соседние блоки могут пересекаться), каждый из которых представлен в результирующем векторе признаков одним значением. Это

значение может быть: 1) средним по блоку; 2) максимальным значением в блоке; 3) главной компонентой в блоке; 4) центральное значение в блоке.

Кроме того, предоставляется возможность выбора главных компонент исходного набора гиперспектральных векторов данных в соответствии с упорядоченной по убыванию последовательностью собственных чисел ковариационной матрицы. Количество выбираемых главных компонент определяется суммарной дисперсией.

К полученному таким образом мультиспектральному образу в полной мере применимы функции системы контролируемой классификации программного комплекса PlanetaMonitoring (включая функцию постклассификации Vote, известную сейчас как мажоритарный фильтр).

Хотя подобное снижение размерности гиперспектральных образов используется достаточно часто (см., например, [4]), применять его следует с осторожностью из-за возможных потерь полезной информации. В частности, метод главных компонент, являясь линейным преобразованием, может разрушать нелинейности, присутствующие в данных [3].

В последнее время в состав системы контролируемой классификации комплекса PlanetaMonitoring нами включен еще один классификатор: классификатор спектрально-углового картирования [5]. Первоначально разработанный специально для классификации гиперспектральных данных этот классификатор не содержит какой-либо специфики этих данных и может быть использован для классификации и многоспектральных данных. В качестве расстояния  $\rho(x, y_i)$  вектора признаков  $x$  до прототипа  $y_i$  класса  $i$  используется угловое расстояние:

$$\rho(x, y_i) = \arccos\left(\frac{x^T y_i}{\|x\| \cdot \|y_i\|}\right).$$

Вектор  $x$  зачисляется в класс, для которого угловое расстояние  $\rho(x, y_i)$  является минимальным. В качестве прототипов классов (ссылочных векторов) авторы алгоритма [5] используют библиотеку лабораторных спектров. Мы в своей реализации алгоритма в качестве прототипов классов используем векторы средних, полученных на основе обучающих и контрольных выборок классов. Для работы с обучающими и контрольными полями используется тот же инструментарий, что и для классификатора Байеса [2]. Классификатор работает в двух режимах: тестовом и рабочем. По результатам тестового режима вычисляется вероятность правильной классификации и строится матрица ошибок. Анализируя соответствующие показатели и выполняя, при необходимости, корректировки обучающих выборок классов, можно добиться необходимого уровня правильной классификации. В рабочем режиме классификатор обрабатывает весь набор данных, используя построенные прототипы классов. При этом может использоваться заданное пороговое значение угла  $\alpha$ : если для вектора признаков  $x$  минимальное значение угла  $\min_i \rho(x, y_i) > \alpha$ , то этот вектор заносится в класс отклоненных векторов.

Для вычислительных экспериментов использовался один из файлов известного набора данных спектрометра AVIRIS [6]. Размер файла 1848\*614, размерность векторов измерений 210. Из этого файла были сформированы два набора векторов признаков размером 1848\*614 векторов: один с размерностью векторов признаков 184 (на основе визуального анализа спектральных составляющих векторов измерений) и второй с размерностью векторов признаков 21 (взяты главные компоненты из каждых 10 последовательных спектральных составляющих). Вычисления производились в рамках модели параллельного программирования OpenMP под управлением ОС Windows-10 на компьютере с 4-ядерным процессором i3-2100. В табл. 1 приведены затраты времени в сек. на обработку указанных наборов данных классификаторами Байеса и спектрально-углового картирования в зависимости от числа выделяемых классов.

Таблица 1:

Размерность векторов признаков	Классификатор Байеса		Спектрально-угловой классификатор	
	15 классов	10 классов	15 классов	10 классов
184	170	105.2	2.4	1.6
21	2.5	1.7	0.35	0.24

На представленных ниже рисунках демонстрируются результаты работы классификаторов. На рис. 1 представлено изображение, полученное совмещением спектральных полос с номерами 30, 1 и 4 исходного

набора данных. Рис. 2 — результат, полученный применением классификатора Байеса к набору векторов признаков размерности 184 (после операции постклассификации). Рис. 3 — результат, полученный применением классификатора спектрально-углового картирования к набору векторов признаков размерности 184 (использовалось пороговое значение  $\alpha = \pi/18$ ). Рис. 4 — результат, полученный применением классификатора Байеса к набору векторов признаков размерности 21 (после операции постклассификации). Белым цветом выделен класс отклоненных векторов. Во всех примерах выделялось по 10 классов.



Рис. 1:



Рис. 2:

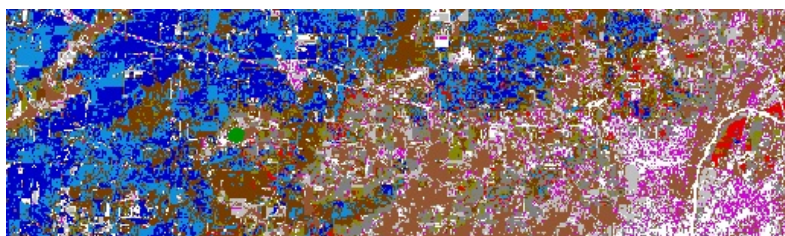


Рис. 3:



Рис. 4:

Включение технологии контролируемой классификации гиперспектральных данных ДЗЗ в состав программного комплекса PlanetaMonitoring позволяет расширить возможности системы контролируемой классификации комплекса по построению карты тематических классов данных ДЗЗ различного спектрального разрешения.

## Список литературы

- [1] Шовенгердт Р.А. Дистанционное зондирование. Модели и методы обработки изображений. Пер. с англ. – М.: Техносфера, 2010. – 557 с.
- [2] Асмус В.В., Бучнев А.А., Пяткин В.П. Контролируемая классификация данных дистанционного зондирования Земли // Автометрия. – 2008. – № 4. – С. 60–67.
- [3] J.P. Marques de Sa. Pattern Recognition. Concepts, Methods and Applications. – Berlin: Springer, 2001. – 318 p.
- [4] П.В. Мельников, И.А. Пестунов, С.А. Рылов. Экспериментальное сравнение методов классификации гиперспектральных изображений высокого пространственного разрешения по спектральным и пространственным признакам. // Региональные проблемы дистанционного зондирования Земли: сб. материалов III Междунар. научной конф., 13–16 сент. 2016 г. – Красноярск : СФУ, 2016. – С. 28–33.
- [5] Kruse, F. A., A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F. H. Goetz The Spectral Image Processing System (SIPS) - interactive viualization and analysis of imaging spectrometer data // Remote Sensing of Environment. – 1993. – № 44. – P. 145–163.
- [6] [https://aviris.jpl.nasa.gov/data/get\\_aviris\\_data.html](https://aviris.jpl.nasa.gov/data/get_aviris_data.html)

*Алексей Александрович Бучнев — к.т.н., старший научный сотрудник Института  
вычислительной математики и математической геофизики СО РАН;  
e-mail: baa@ooi.sscs.ru;*

*Валерий Павлович Пяткин — д.т.н., заведующий лабораторией Института вычислительной  
математики и математической геофизики СО РАН;  
e-mail: pvp@ooi.sscs.ru.*

*Дата поступления — 5 мая 2017 г.*