# Классификация гиперспектральных изображений с использованием кластерной структуры данных

# Е.В. Раменская<sup>1</sup>, М.П. Кузнецов<sup>2</sup>, В.В. Ермаков<sup>1</sup>, О.Р. Баркова<sup>1</sup>, А.А. Бран<sup>1</sup>

<sup>1</sup> Самарский государственный технический университет, Самара, 443100, Россия <sup>2</sup> Московский физико-технический институт, Долгопрудный, Московская обл., 141700, Россия E-mail: iljushinakatja@rambler.ru

Решается задача двухклассовой классификации гиперспектрального изображения, полученного в ходе проведения летных испытаний самолета-лаборатории, с разбиением на кластеры внутри каждого класса. На первом этапе алгоритм выполняет кластеризацию областей интереса, максимизируя функцию правдоподобия для смеси гауссовских распределений. На втором этапе алгоритм использует расстояния Махаланобиса до центров кластеров в качестве признакового пространства и выполняет в этом пространстве классификацию методом дерева решений, минимизирующим ошибку классификации. Алгоритм обладает низкой вычислительной сложностью: обучающая стадия линейно зависит от количества размеченных пикселов. Предложенный метод является интерпретируемым в рамках экспертной области. Подобная классификация позволяет с высокой точностью идентифицировать области интереса на гиперспектральном снимке. Алгоритм позволил отделить нефтезагрязненные участки почвы от других слабоотражающих объектов. Результатом работы алгоритма является изображение, на котором цветом обозначены только участки с присутствием нефтепродукта, а все остальные участки затемнены. Данное изображение имеет то же разрешение, что и исходный снимок. Геопривязка сохраняется, и появляется возможность подсчета окрашенных пикселей изображения на полученном снимке.

**Ключевые слова:** гиперспектральное изображение, кластерная гипотеза, смесь распределений, расстояние Махаланобиса, решающее дерево

Одобрена к печати: 20.11.201 DOI: 10.21046/2070-7401-2017-14-7-9-19

### Введение

Обработка гиперспектральных изображений является одной из развивающихся задач в области дистанционного зондирования. Метод гиперспектрального дистанционного зондирования основывается на регистрировании сигналов с различной длиной волны, отраженных от исследуемого участка поверхности. Таким образом, гиперспектральное изображение (Manolakis, Marden, Shaw, 2003) является трехмерной матрицей, в которой две первые компоненты отвечают за пространственное расположение пиксела, а третья — за его спектральную характеристику. Стандартным примером трехмерного изображения является RGB-изображение, содержащее по третьей компоненте амплитуду сигнала в красном, зеленом и синем цвете. Гиперспектральное изображение отличается существенно большим разнообразием длин волн регистрируемых сигналов: современные гиперспектральные датчики позволяют регистрировать сигналы сотен различных длин волн (Уваров и др., 2014). Например, установленный на спутнике EO-1 гиперспектральный сенсор Hyperion предназначен для получения цифровых изображений земной поверхности в 220 спектральных каналах видимой и инфракрасной области в спектральном диапазоне 400-2500 нм со спектральным разрешением 10 нм. На борту спутника «Ресурс-П» установлена гиперспектральная съемочная аппаратура, позволяющая получать информацию в спектральном диапазоне 400–1100 нм со спектральным разрешением 5-10 нм.

Регистрирование сигналов с различной длиной волны позволяет получать детальную информацию об исследуемом участке земной поверхности, вплоть до его химических свойств. В данной работе решается практическая задача выделения нефтяных загрязнений на участке земной поверхности и задача классификации нефтесодержащих отходов по типам спектров. Нефтесодержащие отходы образуются в процессе добычи, транспортировки и хранения нефти, а также при производстве из нее нефтепродуктов. Для решения этой задачи на изображении экспертом выделяются области нефтезагрязнений, называемые далее областями интереса, а также области, не содержащие нефтезагрязнений. На размеченной выборке строится модель классификации объектов-пикселов (Ermakov, Bogomolov, Bykov, 2012), признаковым описанием которых является спектральная характеристика.

При решении задачи выделения нефтезагрязнений на гиперспектральном изображении ранее были выявлены следующие проблемы. Во-первых, для некоторых видов почв присутствие нефти практически не изменяет интегральный коэффициент отражения, что делает спектр нефти трудно отличимым от темной почвы. В видимом диапазоне спектра отражения (400–750 нм) темный тон почвы может быть обусловлен высоким содержанием влаги, органического вещества или нефтепродуктов (Орлов и др., 1989).

Во-вторых, спектр участка поверхности, регистрируемый датчиком, является наложением спектров различных материалов. Кроме того, на регистрируемый спектр также оказывают влияние положение солнца, погода и наличие в атмосфере выбросов различного состава. Необходимость предобработки данных, например внесение поправки на состав атмосферы, делает затруднительным применение стандартных алгоритмов многомерной регрессии (PLS, PCR).

Из-за перечисленных проблем ранее не удавалось получить классификацию приемлемого качества. При этом при использовании методов неконтролируемой классификации (K-means, Iso-data) (Ball, Hall, 1965) происходило объединение объектов, существенно разных по своей природе, в один класс. Базовые методы контролируемой классификации (метод минимального расстояния, параллелепипедов, максимального правдоподобия) позволили добиться точности распознавания лишь 70–85%, что не является удовлетворительным результатом (Токарева, 2010).

Предлагаемый в данной работе алгоритм классификации гиперспектрального изображения основывается на предположении о кластерной структуре данных. Это предположение заключается в том, что регионы интереса, выделенные экспертом на гиперспектральном изображении, являются объединением областей с малым спектральным расстоянием внутри каждой области.

Чтобы учесть кластерную структуру данных, предлагается следующая схема классификации гиперспектрального изображения, схожая со схемой построения сетей радиальных базисных функций (Broomhead, Lowe, 1988; Orr, 1996). На первом этапе выполняется процедура кластеризации размеченных областей интереса. Для этого выполняется процедура оценки смеси распределений (Dinov, 2008), а также подсчитываются расстояния Махаланобиса (Mahalanobis, 1936) между всеми центрами кластеров. На втором этапе выполняется процедура классификации объектов, признаковым описанием которых является набор расстояний Махаланобиса до вычисленных центров кластеров. В качестве процедуры классификации в данной работе рассматривается алгоритм построения решающего дерева (Quinlan, 1986). Выбор этого простого и универсального метода классификации основан на предположении, что объекты разных классов являются хорошо разделимыми в построенном признаковом пространстве расстояний.

Разрабатываемый метод классификации позволяет добиться высокого качества классификации вследствие использования кластерной структуры данных. Метод является интерпретируемым в контексте моделируемого явления: областями интереса считаются те точки изображения, форма спектров которых схожа с формой спектров центров кластеров. Кроме того, вычислительная сложность процедуры обучения — линейная по количеству объектов, что является необходимым требованием при обработке изображения большого размера, количество точек которого достигает 10<sup>7</sup>.

## Исследуемые данные

Гиперспектральные снимки, используемые в работе, были получены в ходе проведения летных испытаний самолета-лаборатории. Был проведен ряд экспериментальных полетов и съемок территории с использованием фото- и гиперспектральной аппаратуры (Гурьянова и др., 2014). Характеристики использованной аппаратуры и параметров полета приведены в *табл. 1*.

Высота полета, м	700-2630
Скорость, км/ч	125-190
Проекция пикселя, м	0,63-1,28
Ширина полосы захвата, м	510-1920
Число спектральных каналов	3-200
Спектральное разрешение, нм	10-15

Таблица 1. Параметры летных испытаний и характеристики аппаратуры

Пример используемого гиперспектрального изображения, усредненного по спектрам, показан на *puc. 1a*. На некотором наборе снимков задана частичная экспертная разметка исходного изображения, показанная на *puc. 1б*. На изображении выделены регионы интереса — области, спектры которых принадлежат к классу 1. Кроме того, указаны регионы класса 0: достоверно известно, что эти области не являются регионами интереса.

Объектами данного исследования — областями, отнесенными к классу 1, — выступают нефтешламонакопители и илонакопители. Они образованы в ходе эксплуатации установки комплексной подготовки нефти (УКПН). Группа накопителей включает накопители действующие и закрытые на рекультивацию.



Рис. 1. Разметка оригинального изображения: а) исходное изображение; б) экспертная разметка

Характеристика объектов размещения нефтесодержащих отходов представлена в табл. 2.

Наименование объекта	Площадь, м <sup>2</sup>	Объем накоплен- ных отходов, м <sup>3</sup>	Состояние верхнего слоя нефтешламонакопителя
Илонакопитель 1	14100	24 600	Вязкие нефтепродукты
Илонакопитель 3	4900	13700	Застывшие нефтепродукты
Илонакопитель 6	7700	13800	Вязкие нефтепродукты
Шламонакопитель 8	3100	4800	Флотошлам
Шламонакопитель 10	9800	7100	Нефтезагрязненный грунт
Илонакопитель 11	3300	9900	Водонефтяная эмульсия
Илонакопитель 12	3100	6500	Водонефтяная эмульсия
Илонакопитель 13	3200	5500	Водонефтяная эмульсия
Илонакопитель 14	3500	11300	Водонефтяная эмульсия
Илонакопитель 15	3000	4200	Водонефтяная эмульсия
Илонакопитель 16	3100	5900	Водонефтяная эмульсия
Илонакопитель 17	3400	6400	Водонефтяная эмульсия
Илонакопитель 18	3300	5200	Водонефтяная эмульсия
Илонакопитель 19	3200	6100	Водонефтяная эмульсия
Илонакопитель 20	3100	5500	Водонефтяная эмульсия
Илонакопитель 21	3300	5100	Водонефтяная эмульсия
Илонакопитель 22	3000	5700	Водонефтяная эмульсия

Таблица 2. Характеристика объектов исследования

Помимо накопителей в разметку изображения включен аварийный пруд с прилегающими к нему нефтяными ямами (Чертес и др., 2012). Ямы образовались в ходе работы нефтеперерабатывающего завода. Их назначение заключается в размещении плавающего нефтепродукта, собранного в секциях аварийного пруда (три секции). Аварийный пруд выступает в качестве сооружения технологической схемы по промежуточной механической очистке нефтесодержащих сточных вод предприятия. Южнее территории нефтяных ям на расстоянии 160–200 м располагаются иловые карты. Рядом с иловыми картами располагаются шламонакопители.

Секции аварийного пруда имеют следующие размеры: 1 секция — 270×90 м, 2 секция — 55×300 м, 3 секция — 80×140 м. Размер нефтяных ям — примерно 80×80 м. Нефтяные ямы разделены перемычкой (10 м) из ненарушенного природного грунта. Верхним слоем таких объектов является водонефтяная эмульсия.

Иловые карты представляют собой спланированные огражденные участки земли для сушки ила (осадков) очистных сооружений, который образуется на последних этапах очистки. Осадки относятся к IV-V классу опасности для окружающей природной среды.

#### Постановка задачи

Решается задача двухклассовой классификации гиперспектрального изображения с разбиением на кластеры внутри каждого класса. Задано гиперспектральное изображение I, представляемое набором пикселов,  $I = \{x_i\}_{i=1}^N$ , где N — количество пикселов изображения, имеющее порядок  $N \sim 10^7$ , компонента вектора  $x_i$  соответствует частоте электромагнитного спектра, а количество компонент вектора — 36.

Назовем экспертную разметку, набор гиперспектральных пикселов  $x_i$  вместе с метками классов  $y_i \in Y = \{0,1\}$ , обучающей выборкой D в задаче классификации гиперспектрального изображения,  $D = \{(x_i, y_i)\}_{i=1}^m$ , где m — количество размеченных пикселов. Отметим, что количество размеченных пикселов может достигать общего количества пикселов на изображении,  $m \sim N \sim 10^7$ .

Известно, что гиперспектральные данные имеют характерную кластерную структуру. Обозначим набор меток кластеров как  $Y' = \{y'_{01}, ..., y'_{0k_0}, ..., y'_{1k_1}\}$ . Здесь  $y'_{1k}$  — метки кластеров, относящихся к спектрам регионов интереса,  $y'_{0k}$  — метки остальных кластеров. Множество Y' так же как и количество кластеров  $k_0k_1$  является неизвестным. Однако предполагается, что выполнена гипотеза кластерной структуры (*adhoc*): расстояние  $d(x_i, x_k)$  между точками, принадлежащими одному кластеру,  $y'(x_i) = y'(x_k)$ , является достаточно малым. Функция расстояния  $d(x_i, x_k)$  также не является известной: ее выбор должен быть обусловлен лучшим кластерным разделением спектров. В частности, в данной работе рассматривается расстояние Махаланобиса для построения сетей радиальных базисных функций (Broomhead, Lowe, 1988).

Задача заключается в построении классификатора, выполняющего следующие условия:

1. Ошибка классификатора  $S(\cdot)$  на элементах обучающей выборки должна быть минимальной:

$$S(f(x), y) = \sum_{i=1}^{m} [f(x_i) \neq y_i] \rightarrow \min(f).$$
(1)

2. Классификатор *f* должен учитывать неизвестную кластерную структуру данных.

3. Классификатор *f* должен удовлетворять условию масштабируемости, т.е. допускать дообучение при добавлении экспертных данных. При этом при дообучении классификатор должен сохранять свои структурные свойства.

4. Поскольку предполагаемый максимальный размер обучающей выборки составляет порядок  $m \sim 10^7$ , то важным условием является субквадратичность алгоритма обучения по чи-слу элементов выборки.

#### Алгоритм классификации с использованием кластерной структуры

Для построения классификатора f, удовлетворяющего условиям 1–4, используется идея обучения сетей радиальных базисных функций. Функция f строится в виде суперпозиции  $\hat{y}(x) = f(x) = h(d(x))$ , где  $d(\cdot)$  — вектор-функция, осуществляющая построение вторичного признакового пространства, а  $h(\cdot)$  — классификатор во вторичном признаковом пространстве. При этом размерность вектора d(x) является неизвестной оптимизируемой величиной и равна оптимальному количеству кластеров  $k_0 + k_1$ , а каждый элемент вектора d(x) является расстоянием от вектора x до центра соответствующего кластера.

Схема классификации показана на *рис. 2.* На первом этапе для объекта *x* формируется набор признаков  $d_{1k}(x)$ , относящихся к спектрам регионов интереса, и  $d_{0k}(x)$ , относящихся к остальным спектрам; на втором этапе функция  $h: d \rightarrow \hat{y} \in \{0,1\}$  осуществляет классификацию во вторичном признаковом пространстве.



Рис. 2. Схема классификатора

#### Построение смеси распределений для выделения кластеров

Для решения задачи кластеризации и построения вторичного признакового пространства предлагается метод построения гауссовской смеси распределений. Метод построения смеси распределений заключается в моделировании функции правдоподобия классов p(x|y)смесью  $k_y$  компонент,  $ymY{0,1}$ , где компоненты имеют n-мерные гауссовские плотности с некоррелированными признаками:

$$p(x|y) = \sum_{k=1}^{k_y} w_{yk} p_{yk}(x), p_{yk}(x) = N(x; \mu_{yk}, \sum_{yk}), \sum_{k=1}^{k_y} w_{yk} = 1, w_{yk} \ge 0.$$
(2)

В формуле (2) векторы  $\mu_{yk}$  являются средними нормальных распределений и соответствуют центрам кластеров,  $\sum_{yk}$  — матрицы ковариаций, а  $w_{yk}$  — веса компонент смеси. Для оценки средних  $\hat{\mu}_{yk}$  и ковариаций  $\sum_{yk}$  смеси распределений используется EM-алгоритм (Dinov, 2008, Redner et al., 1984).

В качестве вторичного признакового описания спектра рассматривается расстояние Махаланобиса между хи вычисленным центром кластера  $\mu_{vk}$ :

$$d_{yk}(x) = \sqrt{(x - \hat{\mu}_{yk})^T \sum_{yk}^{\infty}} (x - \hat{\mu}_{yk}).$$

# Построение дерева решений для классификации спектров

В качестве функции h(d), осуществляющей классификацию на построенном признаковом пространстве расстояний до центров кластеров, предлагается использовать дерево решений, минимизирующее ошибку классификации (1). Выбор этого алгоритма обусловлен следующими факторами. Во-первых, предполагается, что обучающая выборка спектров является хорошо разделимой в построенном вторичном признаковом пространстве, что верно при выполнении *adhoc* гипотезы кластерной структуры. Это позволяет строить на втором этапе «слабый», но хорошо интерпретируемый алгоритм классификации, каким и является решающее дерево при существенном ограничении его высоты.

Во-вторых, метод C4.5 (Quinlan, 2014), используемый для обучения решающего дерева, требует линейных вычислительных затрат по количеству объектов. Его вычислительная сложность составляет  $O(mk^2)$ , где m — количество объектов, k — количество выделенных кластеров (предполагается, что k не превышает 10). С учетом того, что вычислительная сложность ЕМ-алгоритма для нахождения кластеров составляет O(mki) (*puc. 3*), где n — количество признаков, превышающее 100 для гиперспектральных данных, а I — ограниченное количество итераций, суммарная сложность предлагаемого метода является линейной по количество.



Рис. 3. Классификация гиперспектрального изображения: а) тестовое изображение; б) результат классификации

# Вычислительный эксперимент

В качестве анализируемых данных были использованы данные экспериментальных полетов и съемок территории с использованием фото- и гиперспектральной аппаратуры (Гурьянова и др., 2014). Примеры изображений показаны на *puc. 1a* и *3a*, а экспертной разметки областей интереса — на *puc. 1б*.

На этапе обучения проводилась процедура кластеризации пикселов, относящихся к классу 1. Было выявлено семь оптимальных кластеров, выделенных разными цветами на *рис. 5.* Кроме того, были подсчитаны центры этих кластеров, а также их разброс, определяемый параметрами матрицы Махаланобиса. После этого проводилась процедура обучения решающего дерева в пространстве расстояний до центров кластеров.

После стадии обучения была выполнена процедура классификации полных изображений. Результаты классификации показаны на *рис. Зб* и *5б*. Цвета на изображениях соответствуют цветам кластеров на *рис. 4*.

Алгоритм классификации позволил идентифицировать на гиперспектральном снимке УКПН нефтешламонакопители. На снимках, сделанных в разное время и при разных условиях (состав атмосферы, угол солнца), были определены нефтезагрязнения, различающиеся по своему химическому составу и агрегатному состоянию (водонефтяные эмульсии, застарелые нефтешламы, битумы). Алгоритм позволил отделить нефтезагрязненные участки почвы от других слабоотражающих объектов. Подтверждением работы алгоритма служит также идентифицируемый нефтезагрязненный участок, о существовании которого изначально не было никакой информации. При сопоставлении классифицированного снимка с картой было установлено, что это участок нефтеналивной станции, где вероятность загрязнения почвы достаточно высока.



Рис. 4. Кластеризация спектров



Рис. 5. Классификация гиперспектрального изображения: а) тестовое изображение; б) результат классификации

#### Заключение

В результате данного исследования был разработан метод классификации гиперспектрального изображения с использованием экспертной разметки части снимка. Предлагаемый алгоритм использует предположение о кластерной структуре данных. На первом этапе алгоритм выполняет кластеризацию областей интереса, максимизируя функцию правдоподобия для смеси гауссовских распределений. На втором этапе алгоритм использует расстояния Махаланобиса до центров кластеров в качестве признакового пространства и выполняет в этом пространстве классификацию методом дерева решений.

Благодаря предположению о кластерной структуре данных, разработанный алгоритм позволяет добиться высокой точности классификации. Алгоритм обладает низкой вычислительной сложностью: обучающая стадия линейно зависит от количества размеченных пикселов. Кроме того, предложенный метод является интерпретируемым в рамках экспертной области.

Разработанный метод позволил с высокой точностью идентифицировать на гиперспектральных снимках места нефтезагрязнений и классифицировать загрязнения по химическому составу.

Работа выполнена при финансовой поддержке Минобрнауки РФ (проектная часть госзадания, проект 10.3260.2017/ПЧ).

#### Литература

- 1. Гурьянова А.О., Ермаков В.В., Раменская Е.В., Мандра А.Г. Получение опорных спектральных сигнатур при гиперспектральной съемке // Экология и промышленность России. 2014. № 10. С. 44-47.
- 2. Орлов Д.С., Алмосова Я.М., Бочаринова Е.А., Лопухина О.В. Использование метода отражательной способности нефтезагрязненных почв при дистанционном мониторинге // Аэрокосмические методы в почвоведении. 1989. С. 73-75.
- 3. Токарева О.С. Обработка и интерпретация данных дистанционного зондирования Земли: учеб. пособие: Изд-во Томского политехн. ун-та, 2010. 148 с.
- 4 Уваров И.А., Лупян Е.А., Матвеев А.М., Мазуров А.А., Лаврова О.Ю., Митягина М.И. Организация работы с данными спутниковых гиперспектральных наблюдений для исследования процессов в Мировом океане // Современные проблемы дистанционного зондирования Земли из космоса. 2014. Т. 11. № 1. С. 200–212.
- 5. Чертес К.Л., Зеленцов Д.В., Бальзанников М.И., Гришин Б.М., Андреев С.Ю. Интенсивная биотермическая обработка осадков нефтесодержащих сточных вод // Электрон. науч. журн. Нефтегазовое дело. 2012. № 4. C. 261–266.
- Ball G.H., Hall D.J. Isodata, method of data analysis and pattern classification. Stanford Research Institute Pub-6. lisher, 1965. 76 p.
- 7. Broomhead D.S., Lowe D. Radial basis functions, multi-variable functional interpolation and adaptive networks: Technical report. DTIC Document. Royals Signals and Radar Establishment, 1988. 39 p.
- 8. Dinov I.D. Expectation maximization and mixture modeling tutorial / Statistics Online Computational Resource; University of California, Los Angeles. 2008. 18 p.
- Ermakov V.V., Bogomolov A., Bykov D.E. Oil sludge depository assessment using multivariate data analysis // J. En-9. vironmental Management. 2012. Vol. 105. P. 144-151.
- Mahalanobis P.C. On the generalized distance in statistics // Proc. National Institute of Sciences of India. 1936. Vol. 2. No. 1. P. 49–55.
- Manolakis D., Marden D., Shaw G.A. Hyperspectral image processing for automatic target detection applications // Lincoln Laboratory J. 2003. Vol. 14. No. 1. P. 79–116.
   Orr M.J.L. Introduction to radial basis function networks / Centre for Cognitive Science / University of Edinburgh.
- 1996. URL: https://www.cc.gatech.edu/~isbell/tutorials/rbf-intro.pdf (дата обращения: 12.01.2017).
- 13. Quinlan J.R. Induction of decision trees // Machine learning. 1986. No. 1. P. 81–106.
- 14. Quinlan J.R. C4.5: Programs for Machine Learning. Elsevier, 2014. 302 p.
- 15. Redner R.A. Walker H.F. Mixture Densities, Maximum Likelihood and the Em Algorithm // SIAM Review. 1984. Vol. 26. No. 2. P. 195–239.

# Hyperspectral image classification with use of cluster data structure

E.V. Ramenskaya<sup>1</sup>, M.P. Kuznetsov<sup>2</sup>, V.V. Ermakov<sup>1</sup>, O.R. Barkova<sup>1</sup>, A.A. Bran<sup>1</sup>

<sup>1</sup>Samara State Technical University, Samara 443100, Russia <sup>2</sup>Moscow Institute of Physics and Technology, Dolgoprudny 141700, Moscow Region, Russia E-mail: iljushinakatja@rambler.ru

The problem of two-class classification of a hyperspectral image is solved, with the subdivision into clusters within each class. The images were obtained during flight tests of a laboratory aircraft. At the first stage, the algorithm performs clustering of the areas of interest. The likelihood function for a mixture of Gaussian distributions is maximized. At the second stage, the algorithm uses the Mahalanobis distances to cluster centers as a feature space. In this space, the classification is carried out by the decision tree method which minimizes the classification errors. The algorithm has low computational complexity: the learning stage is linearly dependent on the number of marked pixels. The proposed method is interpreted within the framework of the expert field. The learning phase of the algorithm has linear computational complexity by the number of objects. The proposed method of classification makes it possible to identify with high accuracy the areas of interest in the hyperspectral image. The algorithm is the image in which only areas with the presence of petroleum products are indicated by color, and all other areas are darkened. This image has the same resolution as the original image. Georeference is preserved. It is possible to count the marked pixels of the image in the received image to calculate the area of the contaminated sites.

Keywords: hyperspectral image, cluster hypothesis, mixture of distributions, Mahalanobis distance, decision tree

Accepted: 20.11.2017 DOI: 10.21046/2070-7401-2017-14-7-9-19

# References

- 1. Gur'yanova A.O., Ermakov V.V., Ramenskaya E.V., Mandra A.G., Poluchenie opornykh spektral'nykh signatur pri giperspektral'noi s"emke (Obtaining of Support Spectral Signals during Hyperspectral Survey), *Ekologiya i promyshlennost' Rossii*, 2014, No. 10, pp. 44–47.
- 2. Orlov D.S., Almosova Ya.M., Bocharinova E.A., Lopukhina O.V., Ispol'zovanie metoda otrazhatel'noi sposobnosti neftezagryaznennykh pochv pri distantsionnom monitoringe (The use of oil-contaminated soil reflectance for remote monitoring), *Aerokosmicheskie metody v pochvovedenii*, 1989, No. 1, pp. 73–75.
- 3. Tokareva O.S., *Obrabotka i interpretatsiya dannykh distantsionnogo zondirovaniya Zemli: uchebnoe posobie* (Processing and interpretation of remote sensing data), Izd-vo Tomskogo politekhnicheskogo universiteta, 2010, 148 p.
- 4. Uvarov I.A., Loupian E.A., Matveev A.M., Mazurov A.A., Lavrova O.Yu., Mityagina M.I., Organizatsiya raboty s dannymi sputnikovykh giperspektral'nykh nablyudenii dlya issledovaniya protsessov v Mirovom okeane (Management of hyperspectral remote sensing data for studies of world ocean processes), *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2014, Vol. 11, No. 1, pp. 200–212.
- Chertes K.L., Zelentsov D.V., Bal'zannikov M.I., Grishin B.M., Andreev S.Yu., Intensivnaya biotermicheskaya obrabotka osadkov neftesoderzhashchikh stochnykh vod (Intensive biothermal treatment of precipitation oily wastewater), *Elektronnyi nauchnyi zhurnal Neftegazovoe delo*, 2012, No. 4, pp. 261–266.
- 6. Ball G.H., Hall D.J., *Isodata, a method of data analysis and pattern classification*, Stanford Research Institute Publisher, 1965, 76 p.
- 7. Broomhead D.S., Lowe D., *Radial basis functions, multi-variable functional interpolation and adaptive networks*, Royals Signals and Radar Establishment, 1988, 39 p.
- 8. Dinov I.D., Expectation maximization and mixture modeling tutorial, 2008, 18 p.
- 9. Ermakov V.V., Bogomolov A., Bykov D.E., Oil sludge depository assessment using multivariate data analysis, *Journal of environmental management*, 2012, Vol. 105, pp. 144–151.
- 10. Mahalanobis P.C., On the generalized distance in statistics, *Proceedings of the National Institute of Sciences of India*, 1936, Vol. 2, No. 1, pp. 49–55.
- 11. Manolakis D., Marden D., Shaw G.A., Hyperspectral image processing for automatic target detection applications, *Lincoln Laboratory Journal*, 2003, Vol. 14, No. 1, pp. 79–116.
- 12. Orr M.J.L., *Introduction to radial basis function networks*, URL: https://www.cc.gatech.edu/~isbell/tutorials/rbf-intro.pdf (date of the application: 12.01.2017).
- 13. Quinlan J.R., Induction of decision trees, *Machine learning*, 1986, No. 1, pp. 81–106.
- 14. Quinlan J.R., C4.5: Programs for Machine Learning, Elsevier, 2014.
- 15. Redner R.A., Walker H.F., Mixture densities, maximum likelihood and the em algorithm, *SIAM review*, 1984, Vol. 26, No. 2, pp. 195–239.