

ТЕОРИЯ И МЕТОДЫ ИЗУЧЕНИЯ И ОХРАНЫ ОКРУЖАЮЩЕЙ СРЕДЫ

DOI: 10.36535/0235-5019-2020-12-1

УДК 502/504:001

АЛГОРИТМЫ ОБНАРУЖЕНИЯ, КЛАССИФИКАЦИИ И АНАЛИЗА ИЗОБРАЖЕНИЙ

Д.ф.-м.н., проф. **Мкртчян Ф.А.**

(Институт радиотехники и электроники им. В.А. Котельникова РАН,
Фрязинский филиал, Московская обл., г. Фрязино)

SOME METHODS AND ALGORITHMS FOR DETECTION, CLASSIFICATION AND ANALYSIS OF IMAGES

F.A. Mkrtchyan

В работе анализируются методы и алгоритмы обнаружения, классификации и анализа изображений. Рассматриваются методы кластерного и дискриминантного анализа для обнаружения и классификации явлений на изучаемом пространстве. В качестве информативного параметра для обнаружения исследуется математическая модель "пятнистости" фоновых характеристик изучаемого пространства. Приводится структура и соответствующие модули программного обеспечения. Разработана методика и автоматизированная подсистема качественной интерпретации дистанционных измерений на основе кластерного и дискриминантного анализа.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 16-01-00213а.

In this paper, methods and algorithms for detecting, classifying and analyzing an image are analyzed. The methods of cluster and discriminant analysis for the detection and classification of phenomena in the space under study are considered. As an informative parameter for detection, a mathematical model of the "spotting" of the background characteristics of the space under study is investigated. The structure and corresponding modules of the software are given. The technique and the automated subsystem of qualitative interpretation of remote measurements on the basis of cluster and discriminant analysis are developed.

The study was carried out with the financial support of the Russian Foundation for Basic Research in the framework of the scientific project No. 16-01-00213a.

Введение

Задача классификации земных ландшафтов и акватории заключается в разбиении некоторой группы объектов на классы на основе определенных требований. К одному классу при этом относятся объекты, обладающие объективно общими свойствами (леса, сельскохозяйственные угодья, участки увлажненных почв, океанические поля и т.д.). Для решения этой задачи используются различные алгоритмы теории распознавания образов, математической статистики, кластерного анализа и др.

В настоящее время существует очень много методов распознавания, которые обусловлены в значительной степени многообразием постановок конкретных задач. Задача распознавания заключается в разбиении некоторой группы объектов на классы на основе определенных требований. К одному классу относятся объекты, обладающие объективно общими свойствами. Исходным материалом для решения задачи распознавания являются результаты некоторых наблюдений или непосредственных измерений, называемые первичными признаками.

В выборочном пространстве возможны три ситуации: 1) каждая точка пространства принадлежит всегда только одному определенному классу; 2) некоторые точки принадлежат сразу нескольким классам, и различие состоит только в вероятности этих возможностей; 3) представляющая точка принадлежит то к одному классу, то другому или сразу нескольким. Описанные ситуации соответственно называются детерминированная, вероятностная и индетерминированная [7, 9]. Проверка качества распознавания проводится на контрольной последовательности, состоящей из реализаций, не совпадающих с реализациями обучающей последовательности. При этом в детерминированной и вероятностной ситуациях делаются предположения либо о характере распределения генеральной совокупности, либо о возможной структуре множества обучающих и контрольных последовательностей, либо о типе допустимых правил принятия решений. Математическая задача классификации формулируется с помощью разделяющей функции [9, 13]. Детерминистские алгоритмы основаны на абсолютном доверии к обучающей последовательности. Следствием этого является основное требование к детерминистским алгоритмам: обучающая последовательность должна распознаваться безошибочно. Детерминистскими алгоритмами являются метод потенциальных функций [1], метод дробящихся эталонов [13], метод функции принадлежности [1] и др.

В случае вероятностной ситуации целесообразно рассматривать признаки x_1, x_2, \dots, x_n как случайные величины. Будем считать, что для каждого класса известны многомерная (n -мерная) функция плотности вероятностей вектора признаков x , $f(x/\omega_j)$ и вероятность $P(\omega_j)$ появления изображения j -го класса. Зная эту информацию, необходимо классифицировать поступающие изображения путем минимизации вероятности ошибочного распознавания. В такой постановке задачи распознавание формулируется в виде задачи статистических решений [3, 4, 8, 10, 11, 17, 14, 15].

Один из эффективных методов решения задачи распознавания является метод группового учета аргументов (МГУА) [6] позволяющий строить систему признаков, дающую на контрольной последовательности столь же хорошую разделимость классов, как и на обучающей последовательности. Очень интересной является непрерывно-групповая аппроксимация описаний классов [15], позволяющая получать описания двумерных изображений, инвариантные по отношению к различным геометрическим преобразованиям. Мало изучены задачи распознавания, в которых описания объектов медленно меняются во времени. При этом медленно меняется взаимное расположение классов в пространстве признаков. Желательно иметь алгоритм, позволяющий по мере надобности адаптировать решающую функцию к небольшим изменениям во времени расположения классов. Для решения такой задачи полезно применение идей эволюционного моделирования.

Применение методов кластерного анализа для классификации явлений на изучаемом пространстве

Кластером называется группа точек, связанных каким либо признаком [12]. Буквальный перевод этого слова с английского означает сгущение, скопление, гроздь.

Кластер-анализ относится к классификации при обучении без учителя, то есть когда не имеется обучающей выборки, классификация которой известна заранее.

Задача состоит в разделении набора точек, каждая из которых характеризуется несколькими признаками, на группы так, чтобы члены одной группы были в каком-то смысле более близкими друг к другу, чем к элементам других групп. Информация, используемая в кластер-анализе бывает трех основных типов: многомерные данные, данные о близости и данные о кластерах. Многомерные данные задают значения нескольких переменных для каждой из точек. Данные о близости состоят из значений близости объектов одинакового типа. Близость в разных случаях означает сходство, корреляцию, меру пресечения и т.д. В некоторых случаях субъективные представления о кластерах являются полезной формой данных [12]. Иногда алгоритмы кластеризации используют данные о близости как входные, на выходе же - разбиение на кластеры. В другом подходе начинают с многомерных данных, которые предварительной обработкой преобразуются в данные о близости. Выделяют два типа целей в кластер-анализе: определенные и неопределенные. Определенные цели: например, требуется разделить предприятия по отраслям, или группировка заболеваний по под случаям. При неопределенных целях исследователя обычно интересует просто "посмотреть, что тут такое" или облегчить сравнение и оценку данных, или подготовить проведение дальнейшего анализа. Кластеры называют естественными, если принадлежность к ним естественным образом и достаточно хорошо определяется по данным. Если же в процессе отнесения содержится значительный элемент произвола, то кластеры называются произвольными. Также различными бывают критерии выделения кластеров. Иногда это компактность в признаковом пространстве, иногда более важна четкость их разделения. Могут быть наложены требования на форму кластеров. Следующие различия касаются возможности пересечения кластеров. Пересечение может совсем не допускаться. Может допускаться небольшое пересечение и отнесение объектов на границе к обоим кластерам. В иерархическом кластер-анализе допускается полное включение одного кластера в другой, но не допускаются частичные пересечения. Кроме перечисленных можно выделить и другие характеристики методов кластер-анализа. Существуют накапливающие и разделяющие методы. При накапливающем методе кластеры строятся последовательным добавлением к ним новых точек, при разделяющем кластеры последовательно разбиваются на подкластеры. При иерархическом методе различные кластеры представимы в виде "дерева", тогда как при неиерархическом группы нельзя разбить по каким-то уровням.

Локальные и глобальные критерии. Если функция расстояния в пространстве признаков изменяется от одной части к другой, то процесс анализа - локальный, иначе - глобальный.

Взвешенный и невзвешенный анализ. На различных стадиях метода в котором объекты добавляются к кластерам, различным кластерам или различным направлениям могут придаваться разные веса.

Число кластеров в разных методах либо меняется, либо задается заранее. Метрика, используемая в методах может быть различной. Часто используют евклидово расстояние. Возможны и другие варианты. Часто встречается ситуация, когда собраны большие наборы данных (многомерных), причем структура их неизвестна или, в лучшем случае проведено некоторое начальное предварительное исследование ее. Задачей является разбить n многомерных точек в r -мерном пространстве признаков (измерений) обоснованным, надежным и экономным способом.

Некоторые методы, применяемые для этой цели [12]:

1. Односвязывающий метод (метод ближайшего соседа). Каждая группа вначале состоит из одной точки. Объединение происходит на основе расстояния до ближайшего соседа. Группы, расстояния между которыми минимальны, объединяются. Каждое объединение уменьшает на единицу число групп. Такой метод приводит к кластерам в виде "цепочек".

2. Полносвязывающий метод (самого дальнего соседа). Здесь расстояние определяется как мера удаления самых дальних точек. Расстояние между объединяющими кластерами равно диаметру наименьшей сферы, содержащей их обоих. Этот метод создает компактные кластеры в виде гиперсфер, которые с трудом объединяются с другими кластерами. Каждое объединение уменьшает на единицу число групп.

3. Среднесвязывающий метод (метод Кинга). Расстояние между группами определяется как расстояние между центрами масс. Каждое объединение уменьшает на единицу число групп.

4. Метод k средних. Начинается процедура с k точек, которые выбираются произвольно. В качестве критерия используется минимальное расстояние внутри кластера относительно среднего. После включения нового элемента в кластер, среднее пересматривается. Процесс продолжается до тех пор, пока не размещены все точки.

5. ISODATA. Метод начинается с задания некоторого числа k -количества кластеров. Все элементы относят в соответствии внутрикластерным критерием минимальности. После того как все элементы отнесены, средние пересматриваются, и все проделывается снова до тех пор, пока мы не перестанем получать улучшение внутрикластерного критерия минимальности.

6. Дисперсионный критерий оптимизации. Помещают точки в k кластеров, а затем организуют перераспределение точек в соответствии с дисперсионно-ковариационным критерием. Например, пусть T -матрица, описывающая полный разброс n точек в r -мерном пространстве, W_i - матрица разброса n_i точек i -го кластера, B -матрица разброса центра масс k кластеров. Тогда $T = W + B$, где $W = \sum W_i$. Можно рассмотреть два критерия минимальности: минимизировать след W , или же максимизировать отношение детерминантов $|T| / |W|$. Последний критерий по нескольким причинам предпочтительнее. Так, он инвариантен при линейных преобразованиях.

Среднесвязывающий метод (метод Кинга)

В работе для решения задачи классификации используется среднесвязывающий метод (Кинга), который дает хорошие результаты в тех случаях, когда количество имеющейся информации в предполагаемых кластерах невелико. В соответствии с этим методом (как отмечалось выше) расстояние между группами точек в пространстве признаков определяется как расстояние между центрами масс этих групп. Кластеризация в данном случае основывается на предположении

о том, что близким геофизическим ситуациям соответствуют участки повышенной плотности в пространстве признаков. Среднесвязывающий метод относится к объединяющим процедурам. На начальном этапе всякая точка рассматривается как отдельный кластер. Шаг процедуры состоит в объединении в один двух ближайших между собой кластеров. Каждый кластер характеризуется положением своего центра. Положение нового кластера определяется как место "центра масс" соединившихся кластеров. На каждом шаге можно прекратить действие алгоритма или вывести на печать получившиеся к данному моменту разбиение на кластеры. Можно заранее задавать определенное число, если имеются какие-то предположения в пользу того, что именно такое количество кластеров можно выделить в исследуемом районе. В сравнении с такими объединяющими процедурами как одно- и полносвязывающий методы, данный метод представляется более подходящим для выделения предполагаемых естественных объектов, для которых характерна достаточно сложная форма и в то же время сильная компактность.

Метод дискриминантного анализа

При решении задач распознавания образов в отсутствие априорной информации о параметрах и характере распределений, когда число классов не превышает двух, хорошие результаты дает алгоритм дискриминантного анализа.

Пусть известно, что наблюдаемый МС процесс может либо принадлежать фону С, либо в обследуемом регионе имеется аномалия В. Пусть априори известно, что выборки $\{\|c_{ij}\|\}; i=1, N; j=1, k$ и $\{\|b_{sll}\|\}; s=1, M; l=1, k$ принадлежат классам С и В соответственно. Здесь N и M - объемы выборок; k - размерность пространства признаков, характеризующих измеряемую величину. Решение задачи различения классов С и В может быть осуществлено по методу дискриминантного анализа [9, 13]. Суть этого метода состоит в следующем. В k-мерном пространстве информативных признаков проводится гиперплоскость $\Gamma: a_1x_1 + a_2x_2 + \dots + a_kx_k$ которая делит все пространство на две части. Положение точек совокупностей $\|c_{ij}\|$ с и $\|b_{sll}\|$ относительно Γ определяется $T=(c - \bar{c})^2/D$, где D - выборочная дисперсия величин.

Положение плоскости Γ определяется таким выбором коэффициентов $\{a_i, i=1, k\}$, который обеспечивает максимизацию величины T. В результате для нахождения $\{a_i\}$ необходимо решить систему уравнений

$$\bar{c} = \sum_{i=1}^N c_i; \quad \bar{b} = \sum_{i=1}^M b_i$$

$$c_i = (1 \setminus N) \sum_{j=1}^k a_j c_{ij}; \quad b_i = (1 \setminus M) \sum_{j=1}^k a_j b_{ij}$$

$$\sum_{j=1}^k s_{ij} = \bar{c}_i - \bar{b}_i; (i = \overline{1, k})$$

$$s_{ij} = \sum_{l=1}^N (c_{li} - \bar{c}_i)(c_{lj} - \bar{c}_j) + \sum_{l=1}^M (b_{li} - \bar{b}_i)(b_{lj} - \bar{b}_j)$$

Затем в пространстве информативных признаков строится гиперплоскость Γ_0 , описываемая уравнением $a_0 + a_1x_1 + \dots + a_kx_k$ с $a_0 = (c - b)/2$, и повторяются все предыдущие вычисления с заменой плоскости Γ на Γ_0 . В [3] доказана сходимость этого процесса обучения ЭВМ. Принятие решения об отнесении принятой фиксатором совокупности признаков $d = \{d_1, \dots, d_k\}$ к фоновой или сигнальной осуществляется сопоставлением знака величины $a_0 + a_1d_1 + \dots + a_kd_k$ со знаками c и b . Совпадение знака является критерием принадлежности полученной выборки к тому или другому классу.

Математические модели "пятнистости" фоновых характеристик изучаемого пространства

Особенностью дистанционных измерений является сбор информации, когда на вход обрабатывающей системы поступают данные по маршрутам полета летательного аппарата. В результате получается двумерный образ изучаемого объекта. Одной из моделей этого образа является статистическая модель "пятнистости" изучаемого пространства. В реальных условиях исследование пятен, получение их статистических характеристик и использование их в задаче обнаружения является достаточно сложной проблемой. Необходимо разработать критерии, позволяющие отличать пятна от других явлений. Например, необходимо задание такого порога, превышение которого является признаком пятна. Необходимо также развивать модельное представление процессов обнаружения пятен. Наиболее очевидным и простым способом определения пятен является метод задания порогов. В этом случае к области пятен относится та часть пространства, на которой показатель среды по выбранному каналу превышает или, наоборот, не превышает значения порога. Пусть $y = y(x_1, x_2)$ является функцией координат (x_1, x_2) точек рассматриваемого региона. Если очертить "поверхность уровня" $y = \text{const}$ на поверхности региона, то на нее спроектируются замкнутые кривые уровня y , ограничивающие "пятна". Обозначим их средние диаметры Γ_{yi}^+ ($i=1, n_y$). Соединим центры тяжести площадей ближайших соседних пятен отрезками. Обозначим длины их частей, находящихся вне пятен Γ_{yi}^- ($i=1, n_y$). Будем считать рассматриваемые последовательности реализациями случайного вектора $(\lambda_y^+, \lambda_y^-)$ с двумерной функцией распределения $F_y(\Gamma^+, \Gamma^-)$, зависящей от параметра y , а его называть структурной характеристикой пятнистости параметра y . Эмпирическим аналогом функции $F_y(\Gamma^+, \Gamma^-)$ является двумерная гистограмма $\{m_{y^+}, m_{y^-}\}$ с частотами, обозначающими число пар $(\Gamma_{yi}^+, \Gamma_{yi}^-)$ выборки объема n_y , в точности равных

$$(\Gamma_{yi}^+, \Gamma_{yi}^-) \left(\sum_{\Gamma_{yi}^+, \Gamma_{yi}^-} m_{\Gamma_{yi}^+, \Gamma_{yi}^-} = n_y \right)$$

Если компоненты вектора $(\Gamma_{yi}^+, \Gamma_{yi}^-)$ независимы, то достаточно рассматривать две одномерные гистограммы $\{m_{y^+}\}$ и $\{m_{y^-}\}$ где

$$m_{y^+} = \sum_{\Gamma_{yi}^+} m_{\Gamma_{yi}^+}, \quad m_{y^-} = \sum_{\Gamma_{yi}^-} m_{\Gamma_{yi}^-}$$

при геоинформационном мониторинге необходимы более детальные теоретические и экспериментальные исследования характеристик "пятнистости" исследуемого региона.

Один из вариантов исследования характеристик "пятнистости" вследствие аналитических трудностей можно получить на основе машинного моделирования [2,5, 16]. Алгоритмы имитации изображений "пятнистости" основаны на численном решении алгебраических неравенств, определяющих координаты внутренних точек пятен. Уравнение контуров пятен в общем виде записать невозможно, так как реальные пятна разнообразны по форме и размерам. Поэтому контуры пятен описываются системой простых алгебраических уравнений, связанных между собой соотношениями,

$$\sum_{i=1}^n \varphi_i(x, y) = 0$$

где $\varphi_i(x,y)$ - уравнение элементарной кривой. Для упрощения программной реализации имитации изображений пятнистости в качестве уравнений $\varphi_i(x,y)$ принято уравнение окружности с варьируемыми координатами центра и радиуса. Сложные формы пятен формируются совмещением на плоскости чертежа нескольких окружностей с разными параметрами, что определяется системой неравенств вида:

$$\sum_{i=1}^n \{(x - a_i)^2 + (y - b_i)^2 - r_i^2\} \leq 0$$

где x, y - декартовы координаты внутренних точек пятен, a_i, b_i, r_i - координаты центра и радиус i -ой окружности соответственно n - количество окружностей, составляющих моделируемое изображение. Для имитации случайности фонового распределения пятнистости параметры модели пятнистости a_i, b_i, r_i задаются при помощи датчиков случайных чисел. Меняя законы распределения случайных чисел и их статистические параметры, можно получить разные по статистическим характеристикам изображения пятнистости. Наиболее обоснованным представляется логарифмически нормальное распределение размера пятен, т.е. r_i и пуассоновское распределение на плоскости (x,y) числа пятен, соответствующее экспоненциальному распределению расстояний между проекциями центров пятен на оси x и y .

Структура программного обеспечения

В табл. 1 приведен перечень программных модулей автоматизированной системы классификации явлений на земной поверхности. Важным моментом алгоритмического и программного обеспечения системы является возможность пространственной интерполяции и восстановление данных по трассовым дистанционным и точечным наземным измерениям. Одним из существующих аспектов практической значимости предлагаемой автоматизированной системы является качественная интерпретация и визуализация результатов дистанционных измерений. Программный модуль PENTO осуществляет исследование статистических характеристик "пятнистости". Программный модуль KERPAR реализует алгоритм распознавания образов по методу дискриминантного анализа.

Перечень программных модулей автоматизированной системы классификации явлений на земной поверхности.

Имя программного модуля	Функциональная характеристика модуля
REDUCE	Сокращение экспериментальной информации в памяти ЭВМ.
GARM	Восстановление данных по методу гармонической интерполяции.
SPLINE	Пространственная интерполяция данных по методу сплайнов.
OPTIM	Восстановление данных по методу оптимальной интерполяции.
CLASS1	Кластерный анализ, ориентированный на признаковое пространство дистанционных датчиков
CLASS2	Кластерный анализ, ориентированный на учет локальных отсчетов датчиков.
PENTO	Исследование характеристик "пятнистости".
KERPAR	Реализация алгоритма распознавания по методу дискриминантного анализа.
ARCHIV	Сбор, сортировка и накопление данных наземных и дистанционных измерений.
SPOTS	Картирование результатов обработки на принтере.
SPOTS1	Картирование "пятен" на экране дисплея
TABL	Табличная форма визуализации результатов обработки.

Подсистема исследования характеристик "пятнистости"

Процесс проектирования любой системы включает в себя в качестве важнейших этапов определение потоков данных и графическое описание процессов передачи данных. Потоки данных в системе определяются согласно следующим правилам:

1. Каждому источнику данных соответствует один входной поток.
2. Если имеется совокупность наборов данных, получаемых из нескольких источников, эти наборы распределяются по группам обрабатываемых совместно потоков данных.
3. Если не все потоки данных подвергаются обработке одновременно, процесс обработки делится на этапы, в каждом из которых участвует группа совместно обрабатываемых потоков. Кроме того, должны существовать внутренние потоки данных, связывающие последовательные этапы.
4. Для каждого этапа обработки в системе выделяется основной входной поток, содержащий результаты обработки и дополнительный поток для выдачи оперативных отчетов, сообщений об ошибках и другой вспомогательной информации. Исходя из приведенного определения потока данных, выделены информационные потоки, циркулирующие в подсистеме тематической обработки дистанционных данных о параметрах окружающей среды.

Здесь описывается подблок автоматизированной системы классификации явлений на земной поверхности, реализующий алгоритмы статистического моделирования "пятнистости".

1. Исходные данные, полученные с ИСЗ.
2. Имя устройства, на котором записаны исходные данные.
3. Имя устройства, на которое будут записываться результаты обработки.
4. Имя графического файла.

5. Координаты исследуемого района.
6. Номер канала, по которому производились измерения.
7. Шаг гистограммы.
8. Усреднение.

В результате функционирования системы может быть создано несколько файлов двух типов: графические и файлы данных (символьные). Графические файлы содержат:

1. Координаты района.
 2. Номер канала измерения.
 3. Порог.
 4. Шаг гистограммы.
 5. Сдвиг относительно начала координат.
 6. Графическое изображение гистограмм.
- Символьные файлы содержат следующие данные:

1. Координаты района.
2. Номер канала измерения.
3. Порог.
4. (1+,1-) - характеристики "пятен", а именно: порядковый номер, ширина "пятен", координаты начала и конца "пятен", минимальные, максимальные и средние значения измеряемых величин.

5. Общее количество "пятен", минимальная, максимальная и средняя ширина "пятен".

6. Выборки значений измеряемой величины для 1+ - выбросов и для 1- - выбросов, усреднение и объем выборок (т.е. сумма частот).

7. Гистограммы одномерной и двумерной эмпирической функции распределения значений измеряемой величины.

8. Выборки ширины (1⁺, 1⁻) - выбросов.

9. Гистограммы одномерной и двумерной эмпирической функции распределения ширины выбросов.

10. Выборки средних значений выбросов измеряемой величины.

11. Гистограммы одномерной и двумерной эмпирической функции распределения средних значений.

12. Выборки экстремальных значений выбросов.

13. Гистограммы одномерной и двумерной эмпирической функции распределения экстремальных значений.

Система функционирует в интерактивном режиме. Все данные, за исключением дистанционной информации, вводятся с терминала. Радиометрические данные вводятся с устройства, имя которого указывается оператором (это может быть любое устройство долговременного хранения).

Пороги, для которых строятся гистограммы, вычисляются по формуле:

$$H = T_{\min} + k \frac{T_{\max} - T_{\min}}{d}; (d = 10, k = \overline{1,9})$$

где T_{\min} - минимальное значение измеряемой величины, T_{\max} - максимальное значение измеряемой величины. Эти значения могут, как вводиться с терминала, так и вычисляться автоматически. Первые два и последние два порога не используются.

Имя выходного файла задается один раз для данного района и канала измерений. Таким образом, для каждого района по одному каналу на выходе будет создаваться не менее пяти версий файла с именем. Для каждого порога создается своя версия файла. При желании, некоторые пороги можно исключить из процесса обработки. При изменении координат района или номера канала есть возможность изменить имя выходного файла.

На следующем этапе строятся (I^+ , I^-) - характеристики по данному району и записываются в выходной файл. Затем создаются выборки значений измеряемой величины для I^+ - выбросов и I^- - выбросов и строятся гистограммы эмпирической функции распределения для одномерного и двумерного случая.

Затем можно построить выборки и гистограммы эмпирической функции распределения по следующим характеристикам:

- по ширине выбросов;
- по средним значениям выбросов измеряемой величины;
- по экстремальным значениям выбросов измеряемой величины.

При желании можно отказаться от любой из этих возможностей. Можно изменить координаты района или номер канала и начать процесс сначала.

Подсистема качественной интерпретации и визуализации данных геoinформационного мониторинга

Для выделения характерных зон информации при первичной обработке дистанционных измерений удобно применять среднесвязывающий метод кластерного анализа [5,12], который оказывается эффективным при малых объемах выборок. В рамках автоматизированной системы классификации явлений на земной поверхности реализованы два варианта этого метода в виде программных модулей CLASS1 и CLASS2(см. Табл.1), отличающихся формой организации признаков пространств и алгоритмами сравнения. Программа CLASS1 ориентирована на признаковое пространство отсчетов дистанционных датчиков. Программа CLASS2 выделяет районы одинаковой нестабильности по локальным вариациям отсчетов датчиков. Различие алгоритмов сравнения в этих программах состоит в учете или не учета

взаимосвязи соседних отсчетов датчиков. Программа CLASS1 формирует кластеры без учета географической общности показаний радиометров. Программа CLASS2 формирует непрерывные пространственные кластеры. Программы CLASS1 и CLASS2 автоматически исключают из рассмотрения калибровочную информацию. Программный модуль CLASS1 состоит из 6-ти групп операторов. Первая группа осуществляет организацию массивов данных. Усреднение информации по конечной совокупности отсчетов радиометров производит вторая группа операторов. Третья группа производит поиск минимального межкластерного расстояния, после чего выделяются две точки, на которых это расстояние реализуется. Четвертая группа операторов исследует изменения минимального межкластерного расстояния на данном шаге процедуры и в случае появления резких изменений осуществляет визуализацию структуры кластеров. Пятая группа операторов объединяет ближайшие кластеры и пересчитывает их характеристики. Шестая группа операторов вырабатывает критерий остановки алгоритма.

Задачи обработки дистанционной информации в большинстве случаев требуют оперирования с двумерными изображениями, что предъявляет к программному обеспечению использующихся при этом ЭВМ жесткие требования о наличии соответствующего арсенала процедур. В частности, при анализе изображений в

оптическом или радио диапазонах волн от программного обеспечения требуется выполнение таких операций, как дискретизация непрерывных изображений и представление их в цифровом виде, фильтрация изображений с целью повышения качества вводимой в ЭВМ информации с созданием необходимого уровня контрастности, выделение характерных контуров и т.д.. СВЧ-радиометрическое измерение параметров окружающей среды с целью формирования СВЧ-карт невозможно без средств машинной графики, как на этапе первичной обработки, так и на заключительном этапе визуализации полученных результатов. Поэтому, в рамках автоматизированной системы классификации явлений на земной поверхности была развита методика с соответствующей реализацией на персональном компьютере.

Алгоритмы картирования включают следующие этапы работ:

1. описание массивов данных:
 - а) карты исследуемого региона;
 - б) дистанционные измерения.
2. ввод этих данных.

Программные модули MAP, SPOTS и SPOTS1 осуществляют поиск и визуализацию "пятен" в рассматриваемом регионе на принтере и экране дисплея компьютера соответственно.

Литература

1. Айзерман М.А., Браверман Э.М., Розоноэр Л.И. Метод потенциальных функций в теории обучения машин. М.: Наука, 1970, 383 с.
2. Амбросимов А.К. Автоматизация трансект – анализа при исследовании гидрофизических полей в океане. В сб.: Алгоритмы машинной обработки данных в задачах радиотехники и электроники. М.: ИРЭ АН СССР, 1986, с. 42 –46.
3. Арманд Н.А., Крапивин В.Ф., Мкртчян Ф.А. Методы обработки данных радиофизического исследования окружающей среды. М.: Наука, 1987, 270 с..
4. Вальд А. Последовательный анализ. М.: Физматгиз, 1960, 328 с.
5. Василевич В.И. Статистические методы в геоботанике. М.: Наука, 1969, 232 с.
6. Ивахненко А.Г. Индуктивный метод самоорганизации моделей сложных систем. Киев: Наукова думка, 1982, 296 с.
7. Леман Э. Проверка статистических гипотез. М.: Наука, 1964, 408 с.
8. Мкртчян Ф.А. Обработка данных мониторинга окружающей среды при малых объемах измерений.// Проблемы окружающей среды и природных ресурсов. 1999, № 5, с. 2-15.
9. Мкртчян Ф.А. Оптимальное различение сигналов и проблемы мониторинга. М.: Наука, 1982, 186 с.
10. Мкртчян Ф.А. Эффективность дистанционных мониторинговых систем (ДМС)// Экономика природопользования. № 6., 2014, С. 40-51.
11. Мкртчян Ф.А. Анализ эффективности мониторинговых систем дистанционного зондирования // Экологические системы и приборы, №4, 2017, с. 17-23.
12. Соломон Г. Зависящие от данных методы кластерного анализа. Классификация и кластер. М.: Мир, 1980, с.89 - 97.
13. Турбович И.Т., Гитис В.Г., Маслов В.К. Опознавание образов. М.: Наука, 1971, 246
14. Флейшман Б.С., Мкртчян Ф.А., Крапивин В.Ф. Расчет некоторых пара-

метров мониторинговой системы.// Изв. АН СССР. Техническая кибернетика. 1980, №2, с.208.

15. Фу К. Последовательные методы в распознавании образов и обучении машин. М.: Наука, 1971, 255 с.

16. Krapivin V.F., Varotsos C.A., Soldatov V.Yu. New Ecoinformatics Tools in Environmental Science: Applications and Decision-making. Springer, London, U.K., 2015, P. 903.

17. Mkrtychyan F.A. GIMS-technology for environmental monitoring// Remote Sensing and Modeling of Ecosystems for Sustainability XIII, Proceedings of SPIE, Vol. 9975, No. UNSP 99750F, San Diego , California, USA , 2016.