

# ИСПОЛЬЗОВАНИЕ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ С КРИТЕРИЕМ ДЖЕФРИСА-МАЦУСИТЫ ДЛЯ ВЫБОРА ОПТИМАЛЬНОГО НАБОРА ВХОДНЫХ КАНАЛОВ В ЗАДАЧЕ ДЕШИФРАЦИИ НАЗЕМНЫХ ОБЪЕКТОВ ПО ДАННЫМ ДЗЗ

*А.Ф. Валиуллин*

*Казанский национальный исследовательский технический университет  
им. А.Н. Туполева*

## **Введение**

В современном мире с ростом производства и деятельности человека растет степень воздействия на окружающую среду. Когда происходят техногенные аварии или природные катаклизмы, необходимо незамедлительно получить данные для оценки ситуации. Эти данные необходимы для составления мероприятий, направленных на ликвидацию последствий.

Для получения данных и контроля огромных площадей земной поверхности наиболее эффективными являются методы ДЗЗ, когда по каким-либо причинам наблюдение непосредственно с поверхности Земли не обосновано с экономической точки зрения либо невозможно ввиду ряда причин.

Данные дистанционного зондирования Земли (ДЗЗ), получаемые с оптико-электронных космических аппаратов ДЗЗ, обычно представляют собой цифровые изображения излучения, зарегистрированного в различных спектральных диапазонах. Это могут быть данные как видимого спектра, так и инфракрасного и/или микроволнового диапазонов излучения. Получению, обработке и анализу подобных изображений посвящена обширная литература [1, 2, 3]. Наиболее распространенными сенсорами на данный момент являются различные мультиспектральные спектрометрические приборы, регистрирующие излучение в нескольких (обычно до десяти) спектральных диапазонах (каналах).

При автоматизированной дешифрации таких данных (т. е. построении тематических карт на основе данных ДЗЗ) хорошо зарекомендовали себя различные параметрические и непараметрические методы статистики [1], а также другие методы обработки данных, например нейронные сети (обзор литературы по применению нейронных сетей при анализе данных ДЗЗ представлен в [3]).

В настоящее время начинают получать распространение гиперспектральные спектрометрические приборы, которые позволяют регистрировать от нескольких десятков до нескольких сотен каналов. Например, общедоступные данные EO1/Hyperion [4] представляют собой сцены, состоящие из 220 спектральных полос от 0.4 до 2.5 мкм, со спектральным разрешением от 0.1 до 0.11 мкм. Гиперспектральные спектрометрические приборы призваны увеличить точность и надежность ДЗЗ благодаря съемке в широком радиометрическом разрешении.

Основная проблема обработки гиперспектральной информации заключается в том, что классические методы анализа данных ДЗЗ разрабатывались для применения к мультиспектральным снимкам, т. е. для работы с данными относительно малой размерности. Поэтому использование этих методов для обработки данных существенно большей размерности сопряжено с определенными трудностями. В первую очередь это свя-

зано со сложностью точной оценки параметров моделей на основе обучающих данных. Например, количество обучающих примеров, необходимых при оценке параметров для метода максимального правдоподобия (в предположении о нормальном законе распределения данных), растет как квадрат размерности обрабатываемых данных, а в случае использования непараметрических методов между требуемым объемом обучающей выборки и размерностью данных возникает экспоненциальная зависимость [1].

### 1. Описание работы

Для решения проблемы высокой размерности данных необходима методика, дающая рекомендацию о выборе наилучшей комбинации входных каналов для уверенной дешифрации космических снимков. В работе [5] была описана методика, в которой использовалась имитационная модель, предсказывающая средний уровень излучения для интересующего объекта. Однако в примере (см. рис. 1) видно, что два различных класса объектов могут иметь равные средние. В данном случае модель даст низкую оценку возможности дешифрации на данном наборе исходных каналов при наличии практической возможности разделимости.

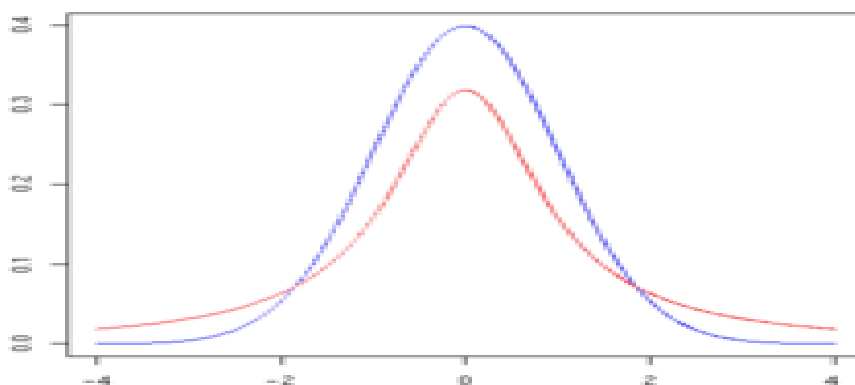


Рис. 1. График плотности распределения двух объектов

Для разрешения этого недостатка необходима оценка, которая будет учитывать расстояние между плотностями вероятности.

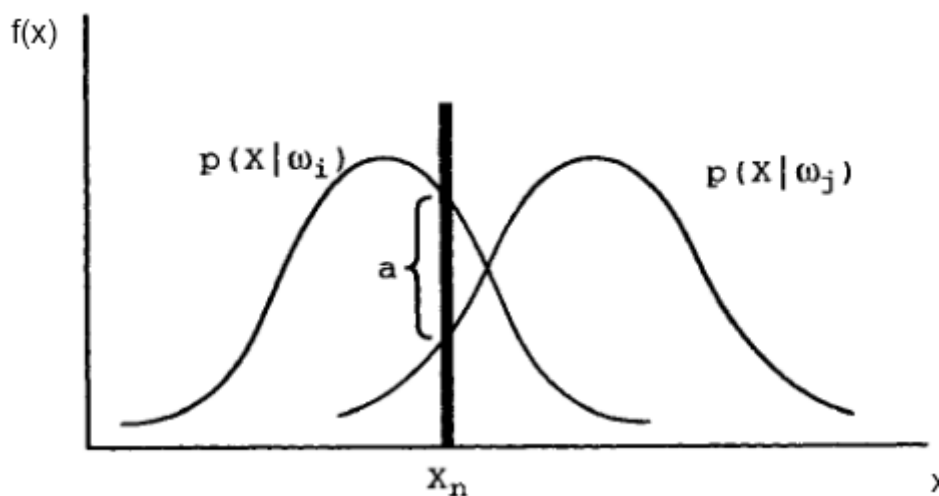


Рис. 2. Геометрический смысл учета расстояния между графиками плотности распределения вероятности

В работе будут использоваться генетические алгоритмы [7]. В качестве фитнес-функции будет выступать оценка Джефриса-Мацуситы [1], которая максимизируется в процессе поиска. На рис. 2 изображена идея использования расстояния между плотностями распределения для получения более объективной оценки комбинаций входных каналов. Генетический алгоритм будет использоваться в качестве механизма эвристического поиска набора входных каналов с максимальным значением оценки разделимости. В качестве исходных данных будут установлены пределы доступных каналов, затем будут рассматриваться закодированные последовательности номеров каналов. Критерием качества найденных решений будет служить расстояние Джефриса-Мацуситы, которое будет рассчитываться для каждого найденного решения. Для проверки адекватности найденной комбинации каналов на них будет обучена нейронная сеть с топологией 2-5-2-1. Критерием качества обучения нейронной сети будет ошибка на тестовом и валидационном множестве.

## 2. Постановка задачи

В работе необходимо найти оптимальный набор входных каналов для наиболее уверенной дешифрации снимков нейронной сетью. Входные каналы получены со спутника MODIS [6] части заповедника «Черные земли», находящиеся в Республике Калмыкия. В наличии 22 канала.

В работе рассматривается задача дешифрации областей, пострадавших от пожара. Также обращается внимание на довольно схожий класс «тени от облаков», т.к. на графике плотности распределения два этих класса имеют примерно равные средние значения. Математическая форма показателя разделимости классов [1] указана ниже:

$$J = \sqrt{\int_{-\infty}^{\infty} \left\{ \sqrt{g(x)} - \sqrt{f(x)} \right\}^2 dx},$$

где  $g(x)$  и  $f(x)$  – функции распределения двух классов.

Однако ввиду отсутствия  $g(x)$  и  $f(x)$  в работе будет использоваться формула для расчета оценки при условии нормального распределения вероятностей рассматриваемых классов:

$$J = \sqrt{2 \left\{ 1 - \frac{\sqrt{|\Sigma_f| |\Sigma_g|}}{\sqrt{2} |\Sigma_f + \Sigma_g|} \exp \left\{ -\frac{1}{8} [\mu_f - \mu_g]^T \left[ \frac{\Sigma_f + \Sigma_g}{2} \right]^{-1} [\mu_f - \mu_g] \right\} \right\}},$$

где  $\Sigma$  – ковариационная матрица,  $\mu$  – средние.

## 3. Результаты

Критерий Джефриса-Мацуситы показал свою эффективность в условиях данной задачи. На рис. 3 визуально видно, что более разнесенные по оси ординат распределения получили более высокую оценку (слева оценка 1.22, справа 0.76). Слева отображены кривые плотности распределения для 8-го и 9-го каналов, где зеленая кривая – плотность распределения сгоревшей территории 9-го канала, синяя кривая – плотность сгоревшей территории 8-го канала, серая кривая – тень от облаков 8-го канала и красная – тень от облаков 9-го канала соответственно. Таким образом получается сравнение двух кривых одного класса обоих каналов с двумя кривыми второго класса обоих каналов.

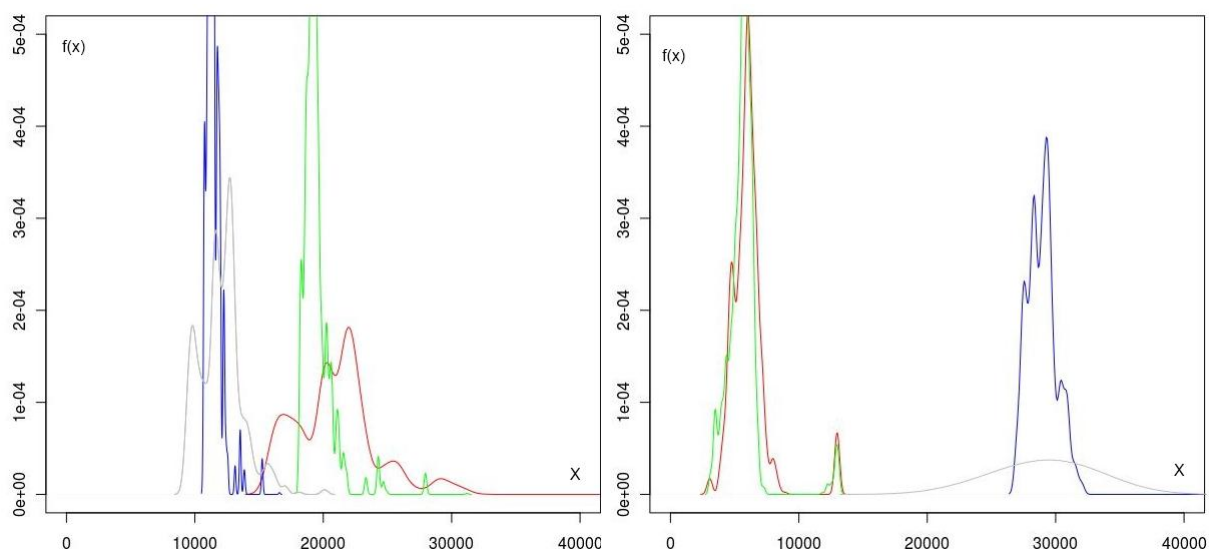


Рис. 3. Графики плотности распределения: зеленый и синий – плотности распределения класса «сгоревшая площадь», красный и серый – «тени от облаков»

Таблица 1. Результаты обучения нейронной сети

1 канал	2 канал	J	Train	Val
<b>8</b>	<b>9</b>	<b>1.221064</b>	<b>1.6915883488419e-06</b>	<b>2.42829360273056e-06</b>
8	10	1.018898	1.69167854095155e-06	0.000881937217142461
1	2	0.9509699	0.012281471583721	0.0161234563422114
5	11	0.7674143	0.025140684272252	0.0218377315961682

Выделенная строка в таблице результатов – это каналы, рекомендуемые генетическим алгоритмом. Видно, что их выбор дает наиболее низкую ошибку как на тестовом, так и на обучающем множестве. Также видно, что при снижении оценки  $J$  падает и качество обучения. Выбор оптимальных каналов занял 2.5 минуты.

Результаты показали, что использование генетических алгоритмов с критерием Джеффриса-Мацуситы дает хорошую рекомендацию при небольших временных затратах.

### Литература

1. Landgrebe D.A. Signal theory methods in multispectral remote sensing. Wiley-Interscience, 2003. 508 p.
2. Richards J.A., Jia X. Remote sensing digital image analysis: an introduction. Springer Verlag, 2006. 454 p.
3. Шовенгердт Р.А. Дистанционное зондирование. Модели и методы обработки изображений. М.: Техносфера, 2010. 582 с.
4. Earth Observing 1 (EO-1) Hyperion – [<http://edcsns17.cr.usgs.gov/eo1/sensors/hyperion>].
5. Валиуллин А.Ф. Имитационная модель системы излучение–датчик для спектрорадиометров дистанционного зондирования Земли // Туполевские чтения XX (принято в печать). Казань, КНИТУ-КАИ.
6. Modis – [<http://modis.gsfc.nasa.gov/>].
7. Рутковская Д., Пилинский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы. 2006. 385 с.