

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

---

# ВОСЬМАЯ СИБИРСКАЯ КОНФЕРЕНЦИЯ ПО ПАРАЛЛЕЛЬНЫМ И ВЫСОКОПРОИЗВОДИТЕЛЬНЫМ ВЫЧИСЛЕНИЯМ

*Томск, 28–30 октября 2015 года*



ИЗДАТЕЛЬСТВО ТОМСКОГО УНИВЕРСИТЕТА  
2015

УДК 519.6

ББК 22.19

В 51

В 51 **Восьмая Сибирская конференция по параллельным и высокопроизводительным вычислениям – Томск:** Изд-во Том. ун-та, 2015. – 164 с.

ISBN 978-5-7511-2389-5

Включены материалы Восьмой Сибирской конференции по параллельным и высокопроизводительным вычислениям, проходившей 28–30 октября 2015 г. в Томском государственном университете при поддержке Министерства образования и науки РФ, Суперкомпьютерного консорциума России, Российского фонда фундаментальных исследований (грант № 15-07-20872) и ЗАО «Intel Software».

Рассмотрены актуальные проблемы организации параллельных вычислений на многопроцессорных системах, современное состояние и перспективы развития методов параллельных вычислений.

Для студентов, аспирантов, преподавателей, научных работников, желающих изучить и практически использовать в научной работе высокопроизводительные вычислительные ресурсы.

**УДК 519.6**

**ББК 22.19**

## СОДЕРЖАНИЕ

Толстых М.А., Фадеев Р.Ю., Мизяк В.Г. Параллельный программный комплекс модели атмосферы для прогноза погоды и моделирования климата .....	5
Шумилов Б.М., Сулайманов З.М. О расщеплении сплайн-вейвлет алгоритмов .....	12
Чуруксаева В.В., Старченко А.В. Численный метод для расчета турбулентного течения со свободной поверхностью .....	20
Ботыгин И.А., Волков Ю.В., Тартаковский В.А. Итеративный параллельный алгоритм формирования климатических кластеров для классификации температурных полей .	26
КупчишинА.Б. Эффективная реализация метода когерентного суммирования и метода шагающих кубов для современных универсальных процессоров.....	31
Колесов А.Е. Численное моделирование задач пороупругости на высокопроизводительных вычислительных системах.....	39
Катаев М.Ю., Лукьянов А.К., Крупский А.С. Параллельные технологии в задачах подготовки многомерных данных к использованию в приложениях.....	48
Катаев М.Ю., Лукьянов А.К., Крупский А.С. Подход к решению задачи кластеризации спутниковых изображений на кластере.....	54
Зюзьков В.М. Новые вероятностные тесты на простоту .....	60
Киреев С.Е., Перепелкин В.А., Ткачёва А.А. Реализация метода IADE-RB-CG в системе фрагментированного программирования LuNA .....	66
Черноскутов М.А. Методы высокопроизводительной обработки графов.....	73
Терентьева М.В., Старченко А.В. Численное решение задачи тепло- и влагообмена в почве .....	80
Эшаров Э.А. Новые подходы к определению серединой линии автомобильной дороги при обработке данных лазерного сканирования .....	87
Потоцкая А.А., Михайлов М.Д. Исследование задачи о разрушении плотины .....	95
Михневич В.Л., Берцун В.Н. Об аддитивной разностной схеме для расчета на кластере теплового состояния многослойного ортотропного цилиндра.....	100
Уkolov P.YU., Merkulova N.N. Исследование модели морфогенеза растений с применением разностных схем .....	109
Федорова О.П. Амшарюк Е.И. К вопросу о выборе словаря признаков при классификации видеоизображений .....	115

Ботыгин И.А. Крутиков В.А. Разработка и реализация методов и алгоритмов построения распределенной инфраструктуры информационно-вычислительных ресурсов для решения крупномасштабных научных задач на основе концепции персонального грида .....	121
Нагиев А.Е. Ботыгин И.А. Программный эксперимент по балансировке нагрузки в распределенной вычислительной системе .....	127
Власенко А.Ю. Гибридный подход к автоматизированной отладке параллельных программ .....	132
Кривошеин С.Г., Окулов Н.Н. Архитектура системы удаленного доступа к вычислительным ресурсам .....	139
Перышкова Е.Н., Мамойленко С.Н., Ефимов А.В. Инструментарий решения масштабируемых задач на распределенных вычислительных системах .....	143
Ерин С.И., Бородина И.А Применение методов фильтрации к данным спутниковых измерений влажности почвы .....	151

Из рисунка 3 видно, что наблюдается рост ингибитора выше значения 1,2 и убывание значений гормона роста до 0,27. Аналогичный характер поведения функций получен с помощью явной схемы и схемы Кранка–Николсон.

Таким образом, в работе проведено исследование модели типа «реакция–диффузия» с применением адаптивных сеток, явной схемы и схемы Кранка–Николсон. Результаты не носят характера окончательных, однако настоящая работа может быть полезна всем, кто занимается изучением математических моделей биофизики и решением задач математической физики.

### Литература

1. Лазарева Г.Г., Миронова В.В., Омелянчук Н.А., Шваб И.В., Вишевков В.А., Горшинченко Д.Н., Николаев С.В., Колчанов Н.А. Математическое моделирование морфогенеза растений // Сиб. журн. вычисл. математики. 2008. Т. 11, №2. С. 151–166.
2. Полякова М.С. О диссипативных структурах в одной модели химической реакции с диффузией // Вестник Московского университета. 1974. №6. С. 643–648.
3. Корн Г. А., Корн Т. Справочник по математике для научных работников и инженеров : определения, теоремы, формулы / под общ. ред. И. Г. Арамановича. СПб.: Лань, 2003. 831 с.
4. Лисейкин В.Д. Обзор методов построения структурных адаптивных сеток // Ж. вычисл. матем. и матем. физ. 1996. Т. 36, № 1. С. 3–41.
5. Дмитриева И.В., Каниметов К.А., Саранча Д.А. Метод подвижных сеток в задаче моделирования миграции леммингов // Численное моделирование в проблеме окружающей среды. Фрунзе: Илим, 1989. С. 109–129.
6. Рихтмайер Р., Мортон К. Разностные методы решения краевых задач. М.: Мир, 1972. 420 с.

## К вопросу о выборе словаря признаков при классификации видеоизображений

DOI 10.17223/978-5-7511-2389-5/17

**О.П. Федорова, Е.И. Амшарюк**

Томский государственный университет, Томск

*opf@math.tsu.ru*

*Исследуются распределения дескрипторов по коллекциям изображений. Проводится сравнительный анализ результатов классификации методами FOREL, k-means и с помощью классификатора Байеса. Предлагается алгоритм отображения наблюдаемого распределения признака на нормальное распределение.*

Ключевые слова: дескрипторы, классификатор Байеса, FOREL, k-means.

Обработка изображения – практически важная и актуальная проблема, включающая задачи классификации и поиска изображений. Такие задачи решаются в медицине, геологии и других областях хозяйственной и научной деятельности при создании и сопровождении баз числовых изображений. Успешное решение задач обработки изображения во многом определяется выбором эффективных признаков (дескрипторов), описывающих изображение, а также используемой моделью данных. Так, при решении задач классификации изображений используются признаки, описывающие цвет, текстуру изображения, ключевые точки.

В настоящей работе использованы следующие базы изображений: мотоциклы (183), подсолнухи (338), гвоздики (398), которые разделены на обучающую выборку и тестовую.

В работе исследуются методы кластеризации FOREL [1] и k-means [2], а также модификация классификатора Байеса для случая нормального распределения признаков [3].

В качестве дескрипторов изображения выбраны: средние по каналам RGB(R, G, B), средние по преобразованию Фурье монохромных изображений по R, G и B (F\_R, F\_G, F\_B), а также моменты Хью [4].

Для каждой базы исследована однородность признаков обучающей и тестовой выборок, в случае, когда выборки имеют нормальное распределение, использовался критерий Стьюдента. Для признаков, распределение которых не является нормальным, использованы критерии знаков и Вилкоксона для связных выборок. Все применяемые тесты показали однородность соответствующих признаков обучающих и тестовых выборок.

Методы FOREL и k-means предварительного обучения не требуют. При кластеризации методами FOREL и k-means применяются обучающие и тестовые выборки отдельно. Ошибка кластеризации методами FOREL и k-means при использовании только средних по каналам RGB очень велика. Использование признаков R, G, B, F\_R, F\_G, F\_B на тестовых базах «мотоциклы–подсолнухи» и «мотоциклы–гвоздики» дает ошибку в пределах 1,3%–1,7%, предварительная стандартизация несколько увеличивает ошибку. Это можно объяснить тем, что текстура изображения при классификации является более значимой характеристикой, чем цвет (рис. 1).

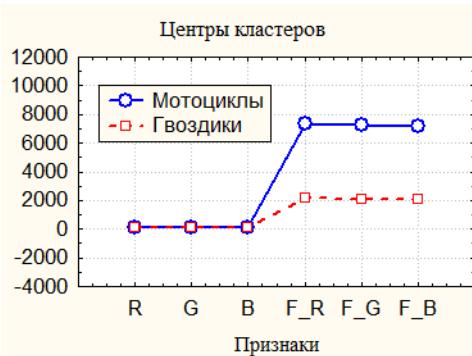


Рис. 1. Центры кластеров в методе k-means

Использование только признаков F\_R, F\_G, F\_B в FOREL и k-means (на тех же парах баз) на уровень ошибки практически не влияет.

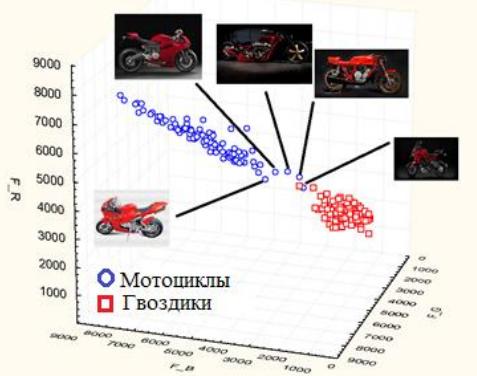


Рис. 2. Диаграмма рассеяния в пространстве признаков

На рисунке 2 показана 3D диаграмма рассеяния, построенная по признакам F\_R, F\_G, F\_B. Диаграмма демонстрирует хорошую разделимость кластеров в пространстве признаков, что объясняет успешную кластеризацию обоими методами.

Использование в качестве признаков инвариантных моментов Хью [4], построенных по монохромным поверхностям R, G и B, при кластеризации баз «мотоциклы–гвоздики» и «мотоциклы–подсолнухи» методами FOREL и k-means получен очень высокий уровень ошибки. Распределение инвариантных моментов для тестовых баз мотоциклов, гвоздик, подсолнухов не является нормальным.

Рассмотрим результаты применения классификатора Байеса. Применение классификатора Байеса требует предварительного обучения. Если используется модификация для нормального распределения признаков и произвольных ковариационных матриц  $\Sigma_1$  и  $\Sigma_2$ , разделяющие функции имеют вид

$$g_i(x) = x' W_i x + w_i^T x + \omega_{i0},$$

где

$$W_i = -\frac{1}{2} \Sigma_i^{-1};$$

$$w_i = \Sigma_i^{-1} \mu_i;$$

$$\omega_{i0} = -\frac{1}{2} \mu_i^T W_i^{-1} \mu_i - \frac{1}{2} \log |\Sigma_i| + \log P(\omega_i),$$

где

$$W_i = -\frac{1}{2} \Sigma_i^{-1}, \quad w_i = \Sigma_i^{-1} \mu_i, \quad \omega_{i0} = -\frac{1}{2} \mu_i^T W_i^{-1} \mu_i - \frac{1}{2} \log |\Sigma_i| + \log P(\omega_i), \quad i = 1, 2.$$

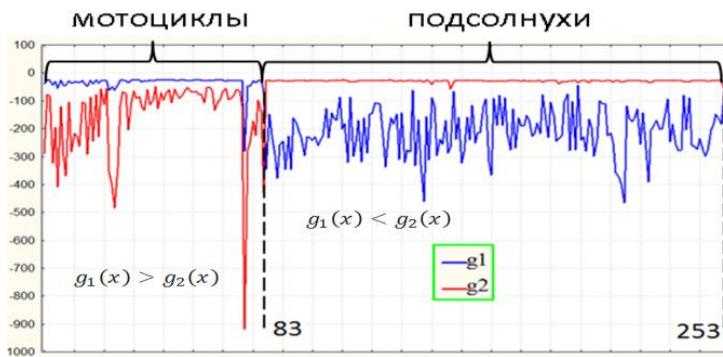


Рис. 3. Разделяющие функции классов «мотоциклы–подсолнухи»

Ковариационные матрицы и векторы средних для каждого класса вычисляются по обучающей выборке пар баз по признакам R, G, B, F\_R, F\_G, F\_B.

На рисунках 3, 4 показаны значения разделяющих функций. Для удобства визуализации в тестовой выборке вначале идут все мотоциклы, а затем подсолнухи (рис. 3) и гвоздики (рис. 4).

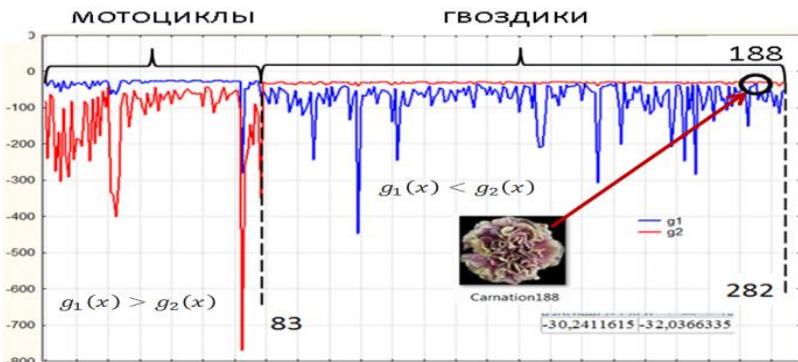


Рис. 4. Разделяющие функции классов «мотоциклы–гвоздики»

В случае классификации мотоциклов и гвоздик только одна гвоздика попадает в чужой класс.

### Отображение наблюдаемого распределения на нормальное

Многие техники статистического анализа разработаны для случая нормального распределения вектора признаков. Представляется интересным построить отображение наблюдаемого распределения на нормальное. Далее представлен такой алгоритм, опирающийся на приближение эмпирического распределения сплайн-функциями.

Рассмотрим разбиение отрезка  $[a,b]$  на  $m$  частей  $x_0 < x_1 < \dots < x_m$ ,  $x_{i+1} = x_i + \Delta_i$ . Запишем сплайн в виде

$$s(x) = \sum_{p=i-1}^{i+2} z_p B_p(x), \quad x \in [x_i, x_{i+1}], \quad i = 0, 1, \dots, m-1.$$

Здесь  $B_i(x)$  – базисные сплайны из  $C_2[a,b]$ , определяемые данным разбиением,  $z_i$  – константы.

Пусть функция  $h(x)$ ,  $x \in [a,b]$ , задана таблицей своих значений на определенном выше разбиении. В работах [5, 6] рассмотрен способ приближения функции сплайнами, сохраняющий интеграл по области задания. Если  $h(x)$  – эмпирическая плотность распределения, то частота попадания значений выборки в промежуток  $[x_i, x_{i+1}]$  равна  $p_i = h\Delta_i$ . Будем задавать коэффициенты сплайна так, чтобы интеграл на каждом промежутке разбиения сохранялся и совпадал с частотой  $p_i$ . В этом случае сплайн можно рассматривать как плот-

ность некоторого распределения, которое приближает функцию  $h(x)$ . Функция распределения такого сплайна

$$S(x) = \int_a^x s(x) dx = \sum_{i=0}^{k-1} p_i + \int_{x_k}^x s(x) dx = z_{k-1} \int_{x_k}^x B_{k-1}(x) dx + \\ + z_k \int_{x_k}^x B_k(x) dx + z_{k+1} \int_{x_k}^x B_{k+1}(x) dx + z_{k+2} \int_{x_k}^x B_{k+2}(x) dx.$$

Для случая равномерной сетки ( $\Delta_i = \Delta$ ) базисные сплайны записываются в виде [7]

$$B_{k-1}(x) = \frac{(1-t)^3}{6}, B_k(x) = \frac{[1+3(1-t)+3t(1-t)^2]}{6}, \\ B_{k+1}(x) = \frac{[1+3t+3t^2(1-t)]}{6}, B_{k+2} = \frac{t^3}{6}, t = \frac{x-x_k}{\Delta}.$$

Алгоритм в этом случае имеет вид:

Шаг 1. По выборке строим сплайн  $y = S(x)$ ;

Шаг 2. Для наблюдаемого значения  $\tilde{x}$  вычисляем  $\tilde{y} = S(\tilde{x})$ ;

Шаг 3. Вычисляем  $z = F^{-1}(\tilde{y})$ ;

Шаг 4. Вычисляем  $x' = \sigma z + \mu$ .

Здесь  $F(x)$  – функция стандартного нормального распределения,  $\mu, \sigma$  – оценки среднего и стандартного отклонения, вычисляемые по исходной выборке.

### Литература

1. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. Новосибирск: ИМ СО РАН, 1999. 270 с.
2. Stuart P. Lloyd. Least squares quantization in pcm // IEEE Transactions on Information Theory, 1982. V. 28, No. 2. P. 129–136.
3. Дуда Р., Харт П. Распознавание образов и анализ сцен. М.: Мир, 1976. 502 с.
4. Hu M.K. Visual Pattern Recognition by Moment Invariants // IEEE Transactions on Information Theory. 1962. V. 8, No. 2. P. 179–187.
5. Федорова О. П. Об одном подходе к приближению функций сплайнами // Вестник ТГУ. Математика и механика. 2008. № 2(3). С. 61–66.
6. Федорова О.П., Кулиш О.В. О задании дополнительных условий в методе построения сплайна, сохраняющего интеграл по области задания // Вестник ТГУ. Математика и механика. 2010. № 1(9). С. 47–52.
7. Завьялов Ю.С., Квасов Б.И., Мирошниченко И.Л. Методы сплайн-функций. М.: Наука, 1980. 352 с.