

РОССИЙСКАЯ АКАДЕМИЯ НАУК  
ОТДЕЛЕНИЕ МАТЕМАТИЧЕСКИХ НАУК РАН  
ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР ИМ. А. А. ДОРОДНИЦЫНА РАН  
УНИВЕРСИТЕТ ИННОПОЛИС  
ПРИ ПОДДЕРЖКЕ  
РОССИЙСКОГО ФОНДА ФУНДАМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ  
КОМПАНИЙ ФОРЕКСИС И ЦСПИР

# Математические методы распознавания образов

## ММРО-16

г. Казань, 6–12 октября 2013 г.

Тезисы докладов 16-й Всероссийской конференции

ТОРУС  
ПРЕСС  МОСКВА  
2013

УДК 004.85+004.89+004.93+519.2+519.25+519.7+519.85

ББК 22.1:32.973.26-018.2

М34

**Математические методы распознавания образов:**  
М34 16-я Всероссийская конференция, г. Казань, 6–12 сентября  
2013 г.: Тезисы докладов. — М.: Торус Пресс, 2013. — 118 с.  
ISBN 978-5-94588-134-1

В сборнике представлены доклады 16-й Всероссийской конференции «Математические методы распознавания образов», проводимой Вычислительным центром им. А. А. Дородницына Российской академии наук и Университетом Иннополис при финансовой и организационной поддержке РФФИ и компаний Форексис и ЦСПиР.

Конференция регулярно проводится один раз в два года, начиная с 1983 г., и является самым представительным российским научным форумом в области распознавания образов и анализа изображений, интеллектуального анализа данных, машинного обучения, обработки сигналов, математических методов прогнозирования.

ББК 22.1:32.973.26-018.2

ISBN 978-5-94588-134-1

© Авторы докладов, 2013

© Вычислительный центр РАН, 2013

## Оргкомитет

**Председатель:** Журавлев Юрий Иванович, *академик РАН*  
Громов Андрей Николаевич  
Инякин Андрей Сергеевич, *к.ф.-м.н.*  
Стрижов Вадим Викторович, *к.ф.-м.н.*  
Тормасов Александр Геннадьевич, *д.ф.-м.н.*  
Чехович Юрий Викторович, *к.ф.-м.н.*

## Программный комитет

**Председатель:** Рудаков Константин Владимирович, *чл.-корр. РАН*  
**Учёный секретарь:** Воронцов Константин Вячеславович, *д.ф.-м.н.*  
Визильтер Юрий Валентинович, *д.ф.-м.н.*  
Горнов Александр Юрьевич, *д.т.н.*  
Дедус Флоренц Федорович, *д.т.н.*  
Донской Владимир Иосифович, *д.ф.-м.н.*  
Дюкова Елена Всеволодовна, *д.ф.-м.н.*  
Кельманов Александр Васильевич, *д.ф.-м.н.*  
Матросов Виктор Леонидович, *академик РАН*  
Местецкий Леонид Моисеевич, *д.т.н.*  
Моттль Вадим Вячеславович, *д.ф.-м.н.*  
Немирко Анатолий Павлович, *д.ф.-м.н.*  
Песков Николай Владимирович, *к.ф.-м.н.*  
Пытьев Юрий Петрович, *д.ф.-м.н.*  
Рейер Иван Александрович, *к.т.н.*  
Рязанов Владимир Васильевич, *д.ф.-м.н.*  
Сойфер Виктор Александрович, *чл.-корр. РАН*  
Устинин Михаил Николаевич, *д.ф.-м.н.*  
Хачай Михаил Юрьевич, *д.ф.-м.н.*

## Краткое оглавление

Композиционные методы анализа данных . . . . .	5
Линейные и метрические методы восстановления зависимостей . .	13
Дискретно-логические методы классификации . . . . .	22
Сложность вычислений и оптимизация . . . . .	31
Обработка изображений . . . . .	43
Анализ изображений . . . . .	51
Анализ видеопоследовательностей . . . . .	56
Распознавание изображений . . . . .	60
Анализ формы изображений . . . . .	68
Приложения: биология и медицина . . . . .	79
Приложения: химия и биоинформатика . . . . .	83
Приложения: анализ текстов и веба . . . . .	87
Прикладные системы . . . . .	95
Содержание . . . . .	106
Авторский указатель . . . . .	115

## Решение задач анализа данных, основанное на линейной комбинации деформаций

*Дьяконов А. Г.*

djakonov@mail.ru

Московский государственный университет имени М.В. Ломоносова;

Вычислительный центр им. А.А. Дородницына РАН

Доклад посвящён теоретическим обоснованиям и практическим приложениям поиска решения задачи регрессии в виде

$$c_1\varphi(B_1) + \dots + c_r\varphi(B_r), \quad (1)$$

где  $B_1, \dots, B_r$  — решения отдельных регрессионных алгоритмов (либо очень простых, либо представимых в виде линейной комбинации «простых»), а  $\varphi$  — «функция деформации», которая выбирается заранее (исходя из специфики задачи) или специально строится.

Поскольку практически любой алгоритм классификации сначала получает некоторые значения, которые естественно назвать «оценками принадлежности к классам», а затем на их основе классифицирует, речь в работе пойдёт не только о регрессии, а о более широком спектре приложений: задачи классификации, рекомендации и т. п. В алгебраическом подходе к распознаванию Ю. И. Журавлёва вводятся операции над алгоритмами, которые, по сути, индуцируются операциями над ответами так называемых распознающих операторов. Поэтому с точки зрения этого подхода выражение (1) определяет вид алгоритма.

Математическое обоснование (1) получено А. Пинкусом:

$$C(\mathbb{R}^n) = \overline{\text{span}} \left\{ \varphi \left( \sum_{i=1}^n a_i x_i + b \right) \mid (a_1, \dots, a_n) \in \mathbb{R}^n, b \in \mathbb{R} \right\},$$

где  $\varphi \in C(\mathbb{R})$  — не полином. Таким образом, любая непрерывная функция может быть приближена двухслойной нейронной сетью, причём в первом слое — фиксированная неполиномиальная функция активации  $\varphi$ , а во втором — тождественная.

В докладе обсуждаются аналогичные результаты в алгебраическом подходе к решению задач классификации. Также приводятся приложения полученных результатов: решения реальных прикладных задач в рамках крупных Международных конкурсов.

Работа поддержана грантом РФФИ № 12-07-00187-а.

- [1] Дьяконов А. Г. Решение задач анализа данных, основанное на линейной комбинации деформаций // Машинное обучение и анализ данных, Москва: ВЦ РАН, 2013. Т. 1, № 5. С. 543–554. <http://jmla.org/papers/doc/2013/no5/Djakonov2013ProblemSolving.pdf>.

## Схема бустинга в задачах коллективного обучения

*Хачай Михаил Юрьевич*<sup>1</sup>

*Поберий Мария Ивановна*<sup>2</sup>

<sup>1</sup>mkhachay@imm.uran.ru, <sup>2</sup>maschas\_briefen@mail.ru

Екатеринбург, Институт математики и механики им. Н. Н. Красовского  
УрО РАН

Системы ограничений, возникающие в задачах принятия решений, оптимизации, распознавания образов и анализа часто являются несовместными, подразумевающими те или иные подходы к их коррекции, связанной с обобщением классического понятия решения. На протяжении последних десятилетий наряду с непрерывным активно развивается дискретный подход к коррекции несовместных систем и несобственных оптимизационных задач, одна из математических формализаций которого связана с исследованием обобщенных решений комитетного типа, порожденных различными логиками голосования. Современная теория комитетных решений и тесно связанных с ними комитетных методов обучения распознаванию опирается на фундаментальные результаты, полученные Вл. Д. Мазуровым.

Построение коллективных обобщенных решений является одним из основных подходов в теории голосования, оптимизации, а также в распознавании образов, где соответствующие алгоритмы получили названия комитетных (committee machines) и бустинга (boosting).

В настоящей работе на примере задачи о минимальном аффинном разделяющем комитете (MASC), в которой используется простейшая логика голосования, основанная на принятии решений большинством голосов, устанавливается близость комитетного подхода и бустинга. Для этого, путем применения традиционного для бустинга игрового подхода к исследованию семейства максимальных по включению отделимых подмножеств, строится полиномиальный приближенный алгоритм для задачи MASC-GP( $n$ ) — подкласса задачи MASC, заданной в пространстве фиксированной размерности  $n > 1$ , в котором разделяемые множества находятся в общем положении, и доказывается, что алгоритм обладает наилучшей на сегодняшний день гарантированной оценкой точности  $O((m/n \ln m)^{1/2})$ , где  $m$  — мощность разделяемых множеств.

Работа поддержана грантами РФФИ № 13-01-00210 и 13-07-00181, грантами 12-01-1016, 12-01-1017/1 Президиума УрО РАН и Молодежным научным грантом ИММ УрО РАН 2013.

- [1] *Хачай М. Ю., Поберий М. И.* Схема бустинга в задачах комбинаторной оптимизации, индуцированных коллективными алгоритмами обучения // Автоматика и телемеханика, Москва: Изд-во, 2014.

## Логические корректоры в задачах классификации по прецедентам

*Дюкова Елена Всеволодовна*<sup>1</sup>

*Любимцева Мария Михайловна*<sup>2</sup>

*Прокофьев Петр Александрович*<sup>3</sup>

<sup>1</sup>edjukova@mail.ru; <sup>2</sup>m.lyubimtseva@gmail.com; <sup>3</sup>p\_prok@mail.ru

Москва, <sup>1,3</sup>ВЦ РАН, <sup>2</sup>ВМК МГУ

Рассматривается подход к задаче распознавания по прецедентам, базирующийся на применении логических и алгебро-логических методов анализа данных. Важнейшими для этого направления исследований являются вопросы построения корректных распознающих алгоритмов.

Классические логические распознающие алгоритмы основаны на поиске в данных конъюнктивных закономерностей или элементарных классификаторов (эл.кл.). В случае целочисленной информации в роли эл. кл. выступают элементарные конъюнкции, определенные на наборах, являющихся признаковыми описаниями объектов. Элементарная конъюнкция считается корректной, если ее интервал истинности не содержит описаний обучающих объектов из разных классов. Корректность распознающего алгоритма обеспечивается корректностью используемых эл. кл.

Представляемый доклад написан по материалам статьи [1] и посвящен исследованию логических корректоров — моделей корректных распознающих алгоритмов, при конструировании которых используются логические и алгебраические методы. Идея синтеза распознающих процедур на основе алгебро-логической коррекции эл. кл. предложена Ю. И. Журавлевым, К. В. Рудаковым и Е. В. Дюковой (1996) и затем развита в ряде работ Ю. И. Журавлева, Е. В. Дюковой, Р. М. Сотнезова и П. А. Прокофьева. В качестве базисных распознающих алгоритмов выступают эл. кл., не обязательно являющиеся корректными, а в качестве корректирующей функции берется булева функция, в частности монотонная. Основной задачей является поиск корректных наборов эл. кл. с высокой распознающей способностью. Для сокращения временных затрат на этапе поиска наиболее информативных корректных наборов эл. кл. используются генетические алгоритмы. Авторами доклада подход продемонстрирован на большом числе реальных задач. В тестировании участвовали различные модели логических корректоров и ряд широко применяемых распознающих процедур.

Работа частично поддержана грантом РФФИ № 13-01-00787-а и грантом президента РФ НШ-4652.2012.1.

- [1] Дюкова Е. В., Любимцева М. М., Прокофьев П. А. Об алгебро-логической коррекции в задачах распознавания по прецедентам // Машинное обучение и анализ данных, 2013. <http://jmla.org/papers/doc/2013>.

## О поточечно корректных операциях над алгоритмами распознавания и прогнозирования

*Шибзухов Заур Мухадинович*

szport@gmail.com

Нальчик, НИИ прикладной математики и автоматизации КБНЦ РАН;

Москва, Центр «Антистихия» МЧС России.

В настоящей работе обсуждается один класс корректных операций над алгоритмами распознавания и прогнозирования, возникающих в связи с проблемой построения корректных алгоритмов.

Корректные операции образуют значительный подкласс корректирующих операций. Они обладают важным свойством – сохраняют свойство корректности алгоритмов.

Типичная задача поиска корректной операции возникает, когда используемый метод обучения позволяет строить множество различных алгоритмов, корректных на своем обучающем множестве и обладающих относительно, но не достаточно хорошим качеством функционирования на контроле.

Основное внимание уделено исследованию поточечно корректных операций над алгоритмами, которые сохраняют свойство поточечной корректности алгоритма, когда алгоритм дает правильные ответы для всех примеров из обучающего множества.

Описана обоснованная схема построения поточечно корректных операций по ответам для алгоритмов регрессии и по оценкам для алгоритмов классификации. Показано также, что поточечно корректные операции можно строить на базе функций типа среднего по Колмогорову и их обобщений.

Работа поддержана грантом РФФИ № 12-01-00162.

- [1] *Шибзухов З. М.* О поточечно корректных операциях над алгоритмами распознавания и прогнозирования // Доклады РАН, М: МАИК Наука, 2013. — Т. 405, № 1, С. 24–27.

## Об агрегированно корректных операциях над алгоритмами

*Матросов Виктор Леонидович*<sup>1</sup>

*Шибзухов Заур Мухадинович*<sup>2</sup>

rector@mpgu.edu<sup>1</sup>, szport@gmail.com<sup>2</sup>

Москва, Московский государственный педагогический университет

В настоящей работе обсуждаются корректные операции над алгоритмами, возникающие в связи с проблемой построения корректных алгоритмов.

Корректные операции образуют значительный подкласс корректирующих операций. Они обладают важным свойством — сохраняют свойство корректности алгоритмов.

Типичная задача поиска корректной операции возникает, когда используемый метод обучения позволяет строить множество различных алгоритмов, корректных на своем обучающем множестве и обладающих относительно, но не достаточно хорошим качеством функционирования на контроле.

Основное внимание уделено исследованию применения агрегирующих функций для построения как поточечно, так и агрегированно корректных операций над алгоритмами.

Описана обоснованная схема построения агрегированно корректных операций на основе агрегирующих функций типа среднего.

Работа поддержана грантом РФФИ № 12-01-00162.

- [1] *Матросов В. Л., Шибзухов З. М.* Об агрегированно корректных операциях над алгоритмами // Преподаватель XXI век, М: Прометей, принято в печать.

## Выбор оптимальных весов в коллективе алгоритмов кластерного анализа

*Бериков Владимир Борисович*

berikov@math.nsc.ru

Новосибирск, Институт математики им. С. Л. Соболева СО РАН,  
Новосибирский государственный университет

Рассматривается задача кластерного анализа: требуется разбить  $N$  объектов, информация о которых имеет форму таблицы «объект–признак», на некоторое число групп в соответствии с заданным критерием однородности. Для решения данной задачи существует большое число методов, отличающихся разными способами понимания однородности, алгоритмами перебора вариантов разбиений и различными ограничениями, позволяющими учитывать специфику конкретной задачи. Поскольку не всегда ясно, какой из методов является наилучшим, а также какие параметры («условия работы» алгоритма) окажутся оптимальными в конкретной ситуации, целесообразно применять коллектив (ансамбль) алгоритмов  $\mu_1, \dots, \mu_M$ , каждый из которых используется для получения  $L_m$  вариантов разбиения,  $m = 1, \dots, M$ . Затем по имеющемуся множеству вариантов разбиений строится консенсусное разбиение.

В докладе обсуждается метод построения коллективного решения с учетом весов алгоритмов, основанный на нахождении матрицы попарных различий между объектами. Элементы матрицы равны

$$h(i, j) = \sum_{m=1}^M \alpha_m(i, j) \frac{1}{L_m} \sum_{l=1}^{L_m} h_{l,m}(i, j),$$

где  $i, j \in \{1, \dots, N\}$  — номера объектов ( $i \neq j$ ),  $\alpha_m(i, j)$  — некоторые неотрицательные веса,  $\sum_m \alpha_m(i, j) = 1$ ;  $h_{l,m}(i, j) = 0$ , если пара  $(i, j)$  объединена алгоритмом  $\mu_m$  в  $l$ -м варианте разбиения в один кластер, и 1 — если в разные кластеры. Для нахождения оптимальных весов, во-первых, необходимо разработать некоторый критерий, а во-вторых, предложить эффективную процедуру их выбора. В докладе обсуждается подход, основанный на минимизации верхней оценки вероятности ошибки классификации, найденной в рамках модели ансамблевого кластерного анализа с латентными классами [1]. Указанная (непосредственно ненаблюдаемая) ошибка оценивается по наблюдаемым характеристикам ансамбля. Разработан соответствующий алгоритм; в докладе приводятся примеры его работы на тестовых и реальных данных.

Работа поддержана грантом РФФИ № 11-07-00346а.

- [1] Бериков В. Б. Коллектив алгоритмов с весами в кластерном анализе разнородных данных // Вестник Томского гос. университета. Управление, вычислительная техника и информатика. 2013. №2(23). С. 22–31.

## Методы агрегирования метрических описаний

*Суворов Михаил Андреевич*<sup>1</sup>

*Майсурадзе Арчил Ивериевич*<sup>2</sup>

<sup>1</sup>suvorov\_m90@mail.ru, <sup>2</sup>maysuradze@cs.msu.ru

Москва, ВМК МГУ

В докладе рассматриваются методы агрегирования метрических описаний. Метрические описания возникают и в теории распознавания, и в прикладных задачах. Многообразие способов сопоставить объекты между собой приводит к большим объемам сравнительной информации. При этом между способами сравнения могут быть зависимости, приводящие к избыточности. Естественно желание избавиться от избыточной информации. Существующие методы агрегирования данных, среди которых в докладе выделены методы матричной факторизации, приспособлены для работы с признаковыми описаниями, но не годятся для метрических описаний.

Введенное понятие метрической конфигурации позволяет сформулировать задачу агрегирования метрических описаний как задачу оптимальной матричной факторизации с дополнительными ограничениями. МГК и НМФ не учитывают эти ограничения, но в докладе показывается, как эти методы могут быть модифицированы для решения сформулированной задачи.

Рассматривается метрический метод главных компонент с минимальными дефектами. Метод, являющийся модификацией метода главных компонент со свободным членом, основан на использовании аддитивной поправки — минимального аддитивного дефекта. Точное значение дефекта известно в редких случаях, что приводит к сложностям, связанным с его оценкой.

В докладе также предлагается модификация неотрицательной матричной факторизации — метрическая неотрицательная матричная факторизация, которая позволяет получать агрегированные метрические описания и избегать трудностей, связанных с использованием минимальных аддитивных дефектов. Доказывается, что сформулированная в терминах матричной факторизации с метрическими ограничениями задача агрегирования метрических описаний может быть решена с сохранением качества аппроксимации на уровне безусловной задачи оптимизации. Показывается, что данное решение может быть конструктивно получено из аналитического решения метода главных компонент.

- [1] *Maysuradze A., Suvorov M.* Methods of metric-based description aggregation based on optimal matrix factorization // Machine Learning and Data Mining in Pattern Recognition, Proceedings Conference MLDM 2013, принята к публикации.

## Агрегирование адаптивных алгоритмов прогнозирования при линейной несимметричной функции потерь

*Романенко А. А.*

alexromsput@gmail.com

Москва, МФТИ

Рассматривается задача прогнозирования временных рядов продаж товаров в розничной сети при линейной несимметричной функции потерь. Функция потерь в *несимметричной линейной задаче* имеет вид

$$\lambda(x, \hat{x}) = \begin{cases} k_1 \cdot |x - \hat{x}|, & x < \hat{x}, \\ k_2 \cdot |x - \hat{x}|, & x \geq \hat{x}, \end{cases} \quad (1)$$

где  $k_1, k_2 \in (0, +\infty)$ .

В качестве базовых алгоритмов предлагается использовать экспоненциальное сглаживание, линейный алгоритм Брауна, Тейла–Вейджа, изменённые для случая несимметричной функции потерь, а также гистограммные алгоритмы прогнозирования, построенные на основе адаптивных алгоритмов прогнозирования.

В качестве алгоритма композиции используется агрегирующий алгоритм В. Вовка. Получена оценка процесса потерь предложенных композиций в худшем случае для функций потерь (1):

$$\mathcal{L}_{AA}(T) \leq (1 + \alpha) \min_{j=1, \dots, M} \mathcal{L}_{B^j}(T) + \frac{1 + \alpha}{\alpha} \frac{(Y_2 - Y_1)K}{4} \ln(M),$$

где  $\mathcal{L}_{B^j}(T)$  — процесс потерь базового алгоритма  $B^j$ ,  $\mathcal{L}_{AA}(T)$  — процесс потерь композиции,  $\alpha \in (0, \infty)$  — параметр композиции,  $Y_1, Y_2 \in \mathbb{R}$  — минимальное и максимальное значение временного ряда,  $K = (2(k_1 + k_2) - k_1 k_2 (4 - k_1 - k_2)) \in (0, +\infty)$ .

В экспериментальной части работы проводилась настройка параметров предложенных алгоритмов на реальных временных рядах продаж товаров в розничной сети. Показано, что для построенных композиций оценки процессов потерь достаточно точны. Предложенные композиции улучшают качество прогнозов базовых алгоритмов на 5%–10%, а также работают лучше ранее известных композиций.

- [1] Романенко А. А. Агрегирование адаптивных алгоритмов прогнозирования при функции потерь с аддитивным членом, зависящим только от величины прогноза // Машинное обучение и анализ данных, — Москва, 2013. <http://jmla.org>.

## Линейные методы восстановления зависимостей по эмпирическим данным в метрических пространствах

*Моттль Вадим Вячеславович*<sup>1</sup>

*Середин Олег Сергеевич*<sup>2</sup>

<sup>1</sup>vmottl@yandex.ru, <sup>2</sup>oseredin@yandex.ru

Тула, <sup>1,2</sup>ТулГУ; Москва, <sup>1</sup>МФТИ, <sup>1</sup>ВЦ РАН

Общепринятый подход к решению задачи восстановления зависимости скрытой характеристики объектов от наблюдаемой характеристики по эмпирическим данным опирается на гипотезу компактности, предполагающую, что объекты, похожие друг на друга в смысле некоторой метрики, принятой наблюдателем, должны обладать близкими значениями скрытой характеристики. В свою очередь, наиболее развитые методы алгоритмической реализации гипотезы компактности основаны на предположении, что объекты являются элементами некоторого линейного пространства, в котором несходство объектов формализовано как естественная метрика. Такие методы, предполагающие погружение множества объектов реального мира в некоторое, вообще говоря, гипотетическое линейное пространство, мы будем называть линейными.

Основная математическая идея линейного подхода к восстановлению зависимостей заключается в погружении произвольного множества объектов реального мира, рассматриваемого как метрическое пространство, в линейное пространство с так называемым индефинитным скалярным произведением, отличающимся от обычного скалярного произведения только отсутствием неотрицательной определенности симметрических матриц значений для конечных совокупностей элементов, и представимого как результат вычитания друг из друга двух обычных скалярных произведений. Линейные пространства с индефинитным скалярным произведением называют псевдоевклидовыми пространствами. В частности, если исходная метрика обладает некоторыми особенно «удобными» свойствами, то получается обычное евклидово пространство.

В докладе показано единство и различие таких общепринятых задач восстановления зависимостей как построение регрессионных моделей и обучение распознаванию объектов двух классов.

Работа поддержана грантом РФФИ № 11-07-00409.

- [1] *Моттль В. В.* Метрические пространства, допускающие введение линейных операций и скалярного произведения // Доклады Российской академии наук, 2003. — Т. 388, № 3. — С. 1–4.
- [2] *Абрамов В. И., Середин О. С., Моттль В. В.* Обучение распознаванию образов по методу опорных объектов в евклидовых метрических пространствах с аффинными операциями // Известия ТулГУ, Естественные науки, Вып. 2. — Тула, 2013.

## Метрическая коррекция матриц парных сравнений

*Двоенко Сергей Данилович*<sup>1</sup>

*Пшеничный Денис Олегович*<sup>2</sup>

<sup>1</sup>dsd@tsu.tula.ru, <sup>2</sup>denispshenichny@yandex.com

Тула, <sup>1,2</sup>Тульский государственный университет

В задачах интеллектуального анализа экспериментальные данные часто сразу представлены результатами парных сравнений объектов между собой.

В работе рассматривается подход к регулируемой коррекции матриц парных сравнений с целью устранения метрических нарушений. Предполагается, что элементы множества, представленные матрицей парных сравнений, требуется погрузить в некоторое метрическое пространство, например, евклидово.

Показано, что метрические нарушения возникают не только на тройках, но и на подмножествах, содержащих больше элементов. Такие нарушения во взаимном расположении элементов относительно друг друга также приводят к неположительной определенности матрицы скалярных произведений, даже если на всех тройках элементов метрических нарушений нет.

Построены алгоритмы коррекции нормированных и ненормированных матриц парных близостей, интерпретируемых как скалярные произведения, и матриц парных различий, интерпретируемых как расстояния.

Показано, что суть нарушений, как на тройках, так и на большем числе элементов множества, заключается, в итоге, в нарушении теоремы косинусов. Это удастся показать в явном виде только для нормированной матрицы парных сравнений, имеющих смысл скалярных произведений.

В этих условиях потребовалась разработка специальных процедур для корректировки произвольных матриц парных сравнений, которые могут быть матрицами расстояний и ненормированными матрицами близостей или скалярных произведений.

Работа поддержана грантом РФФИ № 13-07-00010.

- [1] Двоенко С. Д. Пшеничный Д. О. О метрической коррекции матриц парных сравнений // «Машинное обучение и анализ данных», Москва: ВЦ РАН, 2013. <http://jmla.org>.

## О локализации отрицательных собственных значений

*Двоенко Сергей Данилович*<sup>1</sup>

*Пшеничный Денис Олегович*<sup>2</sup>

<sup>1</sup>dsd@tsu.tula.ru, <sup>2</sup>denispshenichny@yandex.com

Тула, <sup>1,2</sup>Тульский государственный университет

В отсутствие исходного признакового пространства условием корректного погружения данного множества объектов в метрическое пространство является неотрицательная определенность матрицы парных сравнений элементов множества друг с другом. В этом случае близости интерпретируются как скалярные произведения, а соответствующие различия — как расстояния.

Согласно закону инерции квадратичных форм известно, что число смен знаков при просмотре главных миноров матрицы скалярных произведений совпадает с числом ее отрицательных собственных значений.

Предлагается определить такой порядок просмотра элементов множества, чтобы смены знаков главных миноров матрицы парных сравнений ранга  $n$ , имеющей  $v$  отрицательных собственных значений, происходили преимущественно в конце последовательности. В идеальном случае соответствующая перестановка элементов множества определит такой порядок просмотра его элементов, при котором главный минор на шаге  $n - v + 1$  впервые окажется отрицательным, а знаки последующих  $v - 1$  миноров будут чередоваться.

В этом случае нужно будет скорректировать парные сравнения не более, чем  $v$  объектов. В этом смысле отрицательные собственные значения в неположительно определенной матрице парных сравнений окажутся локализованными.

Таким образом, группировка в конце последовательности объектов, предположительно вносящих метрические искажения, позволяет уменьшить дополнительные нарушения, которые могут возникнуть от других объектов в процессе регулируемой коррекции.

Работа поддержана грантом РФФИ № 13-07-00010.

- [1] Двоенко С. Д., Пшеничный Д. О. О метрической коррекции матриц парных сравнений // «Машинное обучение и анализ данных», Москва: ВЦ РАН, 2013. <http://jmla.org>.

## Поиск наибольшей гиперполости в линейном пространстве данных\*

*Губарева Анна Александровна*<sup>1</sup>

*Ларин Александр Олегович*<sup>2</sup>

*Середин Олег Сергеевич*<sup>3</sup>

*Сулимова Валентина Вячеславовна*<sup>4</sup>

<sup>1</sup>a.a.gubareva@yandex.ru, <sup>2</sup>ekzebox@gmail.com,

<sup>3</sup>oseredin@yandex.ru, <sup>4</sup>vsulimova@yandex.ru

Тула, <sup>1,3,4</sup> Тульский государственный университет

Москва, <sup>2</sup> Московский физико-технический институт

Задача поиска наибольшей гиперполости в линейном пространстве данных может быть отнесена к классу задач описания данных. Необходимость построения такого описания может потребоваться, например, при исследовании свойств пространства, в которое отображаются данные. В частности, ярким примером такой ситуации может служить исследование эволюционного пространства некоторого гена в метагеномике.

Под эволюционным пространством гена в данном случае понимается пространство, в котором все возможные варианты гена (существовавшие, существующие или гипотетически возможные, но еще не реализовавшиеся) представлены точками, а расстояния между объектами соответствуют эволюционным расстояниям между вариантами.

Обнаружение гиперполости в эволюционном пространстве позволило бы определить место локализации общего предка существующих вариантов исследуемого гена и подтвердило бы предположения современных микробиологов о ходе глобальной эволюции.

В данной работе поставлена и решена задача поиска в линейном пространстве наибольшей гиперполости в виде гиперсферы максимального радиуса, не содержащей данные. Данная задача сформулирована как многоэкстремальная задача оптимизации с ограничениями и выражена в терминах признакового пространства и в терминах потенциальных функций. Предложенный подход наследует от одноклассового метода опорных векторов форму представления центра гиперсферы в виде линейной комбинации небольшого количества так называемых опорных объектов. Экспериментальные исследования на модельных данных демонстрируют корректность решения поставленной задачи.

Работа поддержана грантом РФФИ № 11-07-00728а.

- [1] *Губарева А. А.* Поиск наибольшей гиперполости в линейном пространстве данных // Известия ТулГУ. Технические науки, №10, 2013.

---

Работа выполнена при финансовой поддержке РФФИ, проект № 11-07-00728а.

Всероссийская конференция ММРО-16, г. Казань, 6–12 октября 2013 г.

## Логистическая регрессия в метрических пространствах

*Абрамов Вадим Игоревич*<sup>1</sup>

*Середин Олег Сергеевич*<sup>2</sup>

*Моттль Вадим Вячеславович*<sup>3</sup>

<sup>1</sup>v.abramov.work@gmail.com, <sup>2</sup>oseredin@yandex.ru,

<sup>3</sup>vmottl@yandex.ru

Москва, <sup>1,3</sup>Московский физико-технический университет

Тула, <sup>2,3</sup>Тульский государственный университет

Москва, <sup>3</sup>Вычислительный центр им. А. А. Дородницына РАН

В работе предложен метод беспризнакового обучения распознаванию образов в множествах объектов, представленных только некоторой числовой функцией парного отношения между ними, обладающей свойствами метрики специального вида – евклидовой метрики. Заявленный метод практически полностью аналогичен методу логистической регрессии, но использует евклидову метрику между парами объектов для нахождения расстояния от объекта до дискриминантной гиперплоскости. Переход к использованию евклидовой метрики для описания объектов становится возможным за счет введения понятия диполя и соответствующих ему аффинных операций на множествах объектов.

Преимущество использования евклидовой метрики в качестве исходной функции попарного представления объектов вместо потенциальной функции заключается в том, что все линейные методы обучения распознаванию образов инвариантны к выбору нулевого элемента линейного пространства, а использование зависящих от нуля потенциальных функций вносит избыточность в описание объектов.

Статья является пятой в цикле систематических исследований так называемого беспризнакового подхода применительно к классическим линейным методам и основывается на многих результатах исследования этого подхода в отношении метода опорных векторов.

Работа поддержана грантом РФФИ № 11-07-00409.

- [1] *Абрамов В.И., Середин О.С., Моттль В.В.* Обучение распознаванию образов по методу опорных объектов в евклидовых метрических пространствах с аффинными операциями // Известия ТулГУ, Естественные науки, Вып. 2, Тула, 2013.

## Байесовская логистическая регрессия в задаче обучения распознаванию образов при смещении решающего правила

*Турков Павел Анатольевич*<sup>1</sup>

*Красоткина Ольга Вячеславовна*<sup>2</sup>

*Моттль Вадим Вячеславович*<sup>3</sup>

Тула, <sup>1,2,3</sup>Тульский государственный университет

Москва, <sup>2,3</sup>Московский физико-технический университет

Москва, <sup>2,3</sup>Вычислительный центр им. А. А. Дородницына РАН

Под потоком данных понимается упорядоченная последовательность объектов, поступающих так часто, что не имеется возможности хранить их в памяти. Это приводит к необходимости использования специальных online-методов обучения. Кроме того, с течением времени свойства исследуемого явления могут изменяться, что означает наличие так называемого «смещения концепта» (concept drift).

В данной работе представлен механизм обучения распознаванию образов для задач со смещением концепта, данные в которых поступают в виде потока. Указанный механизм построен на использовании Байесовской логистической регрессии. Полученный алгоритм обучения распознаванию образов имеет вычислительную сложность, пропорциональную длине обучающей выборки.

Работа поддержана грантом РФФИ № 12-01-31524.

- [1] *Красоткина О. В., Турков П. А., Моттль В. В.* Байесовская логистическая регрессия в задаче обучения распознаванию образов при смещении решающего правила // Известия Тульского государственного университета. Технические науки, выпуск 2, 2013, С. 177–187.

## Применение Машины Релевантных Объектов в задачах восстановления числовых зависимостей

*Разин Николай Алексеевич*<sup>1</sup>

*Черноусова Елена Олеговна*<sup>2</sup>

*Красоткина Ольга Вячеславовна*<sup>3</sup>

*Моттль Вадим Вячеславович*<sup>4</sup>

<sup>1</sup>nrmanutd@gmail.com, <sup>2</sup>elchernousova@inbox.ru,

<sup>3</sup>o.v.krasotkina@yandex.ru, <sup>4</sup>vmottl@yandex.ru

Москва, <sup>1,2,3,4</sup>Московский физико-технический университет

Москва, <sup>3,4</sup>Вычислительный центр им. А. А. Дородницына РАН

Тула, <sup>3,4</sup>Тульский государственный университет

В работе рассматривается задача беспризнакового распознавания образов в предположении, что объекты попарно сравниваются при помощи произвольной действительной функции. Такой подход является гораздо более общим, чем традиционный метод потенциальных функций (кernels), требующий положительной полуопределенности матрицы функции сравнения объектов. Последнее требование в большинстве случаев является чрезмерным, причем обучение еще более усложняется, если существует несколько различных способов сравнительного представления объектов. В таких случаях экспериментатор вынужден решать задачу исключения как избыточных базисных объектов для сравнительного представления объектов обучающей совокупности, так и способов сравнения. В терминах общего пространства попарного сравнительного представления объектов предлагаемая постановка становится математически аналогичной классической задаче отбора признаков. Получившийся выпуклый критерий обучения аналогичен методу релевантных векторов Типпинга, но является существенно более общим, поскольку содержит структурный параметр, контролирующий селективность отбора.

- [1] *Разин Н. А., Черноусова Е. О., Красоткина О. В., Моттль В. В.* Применение Машины Релевантных Объектов в задачах восстановления числовых зависимостей // Машинное обучение и анализ данных, 2013. — Т.1, №5, С. 574–585. <http://jmla.org/papers/doc/2013/no6/Author2013Title.pdf>.

## Численная реализация алгоритмов селективного комбинирования разнородных представлений объектов в задачах распознавания образов

*Разин Николай Алексеевич*<sup>1</sup>

*Моттль Вадим Вячеславович*<sup>2</sup>

nrmanutd@gmail.com<sup>1</sup>, vmottl@yandex.ru<sup>2</sup>

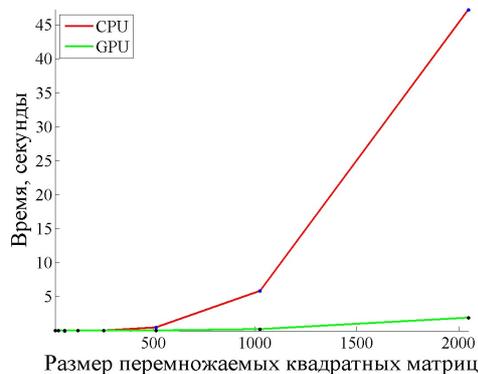
Москва, <sup>1,2</sup>Московский физико-технический университет

Москва, <sup>2</sup>Вычислительный центр им. А. А. Дородницына РАН

Тула, <sup>2</sup>Тульский государственный университет

В работе рассматриваются методы решения задач беспризнакового распознавания образов в предположении, что объекты попарно сравниваются при помощи произвольной действительной функции. Такой подход является гораздо более общим, чем традиционный метод потенциальных функций (кernels), требующий положительной полуопределенности матрицы функции сравнения объектов.

Важное преимущество предлагаемых алгоритмов перед существующими заключается в том, что они хорошо распараллеливаются на современных многопроцессорных вычислительных системах, что позволяет использовать мощные кластеры для быстрого решения задач с большим объёмом данных. Полученное ускорение по сравнению с наивными реализациями алгоритмов умножения матриц и решения систем линейных уравнений доходит до 25 раз (см. Рис. 2) на обычной видеокарте.



- [1] *Разин Н. А., Моттль В. В.* Численная реализация алгоритмов селективного комбинирования разнородных представлений объектов в задачах распознавания образов // Машинное обучение и анализ данных.

## Беспереборный метод скользящего контроля для верификации квадратично-модульной регрессионной модели

*Красоткина Ольга Вячеславовна*<sup>1</sup>

*Моттль Вадим Вячеславович*<sup>2</sup>

*Разин Николай Алексеевич*<sup>3</sup>

*Черноусова Елена Олеговна*<sup>4</sup>

<sup>1</sup>*o.v.krasotkina@yandex.ru*, <sup>2</sup>*vmottl@yandex.ru*,

<sup>3</sup>*nrmanudt@gmail.com*, <sup>4</sup>*lena-ezhova@rambler.ru*

Москва, <sup>1,2,3,4</sup>Московский физико-технический университет

Москва, <sup>1,2</sup>Вычислительный центр им. А. А. Дородницына РАН

Тула, <sup>1,2</sup>Тульский государственный университет

В докладе рассматривается оценивание линейной регрессионной зависимости с помощью критерия, штрафующего сложность модели двумя способами: в виде суммы квадратов и в виде суммы модулей компонент вектора параметров. Критерий с описанной квадратично-модульной регуляризацией в англоязычной терминологии получил название Elastic Net. Вклад в критерий каждого из штрафных слагаемых регулируется двумя структурными параметрами. Как правило, выбор этих параметров осуществляется эмпирическими методами оценивания обобщающей способности модели, типа ресэмплинга, что неизбежно сопряжено с вычислительной сложностью таких методов. В докладе предлагается беспереборный способ вычисления критерия скользящего контроля (в англоязычной терминологии — *leave-one-out*), что позволяет выбрать структурные параметры, избежав трудоемких вычислений. Основная идея предлагаемого метода основана на рассмотрении разбиения множества признаков в точке минимума критерия Elastic Net, как вторичного нечислового параметра регуляризации.

Работа поддержана грантом РФФИ № 11-07-00409.

- [1] *Красоткина О. В., Моттль В. В., Разин Н. А., Черноусова Е. О.* Беспереборная кросс-валидация отбора признаков в линейной регрессионной модели // Известия ТулГУ, Технические науки, Вып. 4, Тула, 2013.

## Исследование погрешности оценок скользящего экзамена

*Неделько Виктор Михайлович*

nedelko@math.nsc.ru

Новосибирск, Институт математики СО РАН

В работе на модельных задачах проводится сравнение различных вариантов оценки скользящего экзамена, таких как leave-one-out и  $K$ -fold cross-validation, а также оценки, основанной на эмпирическом риске с поправкой на смещение. В качестве методов классификации рассмотрены дискриминант Фишера и гистограммный классификатор.

В литературе можно встретить утверждение, что наилучшим выбором является использование  $K$  от 5 до 10, и что такой вариант предпочтительнее чем leave-one-out.

Действительно, есть ряд доводов в пользу такого утверждения.

Первый довод состоит в том, что leave-one-out менее устойчива. Легко привести примеры выборок для случая двух классов, где эта оценка равна 1. Однако при малых  $K$  оценка cross-validation становится существенно смещённой (поскольку строится при меньшем объёме обучающей подвыборки).

Другой довод состоит в том, что при  $K \ll N$ , где  $N$  — объём выборки, применима классическая оценка доверительного интервала для риска. Очевидно, что оценка  $K$ -fold cross-validation не хуже, чем оценка по одной контрольной выборке объёма  $\frac{N}{K}$ . Фактически же оценка cross-validation существенно лучше последней, но нет приемлемых аналитических оценок, насколько лучше. Для leave-one-out вообще неизвестно приемлемых оценок точности, но то, что мы не можем оценить её точность, не означает, что она менее точна, чем cross-validation.

Исследование, проведённое в данной работе, свидетельствует о практической равноценности оценок leave-one-out и 5-fold cross-validation. Хотя первая оценка в большинстве испытаний оказывается слегка точнее, в ряде случаев она даёт существенно большую погрешность. На практике этот эффект, однако, вряд ли является существенным.

Также можно сделать вывод о перспективности оценки на основе эмпирического риска с поправкой на смещение. Данная оценка имеет сравнимую точность, но гораздо меньшую трудоёмкость в вычислительном плане.

Работа поддержана грантом РФФИ № 11-07-00346-а.

- [1] Неделько В. М. Исследование погрешности оценок скользящего экзамена // Машинное обучение и анализ данных, Москва, 2013. — С. 5–25. [http://jmla.org/papers/doc/2013/no6/jmla\\_Nedelko.pdf](http://jmla.org/papers/doc/2013/no6/jmla_Nedelko.pdf).

## Комбинаторные оценки переобучения пороговых решающих правил

*Ишкина Шаура Хабировна*

*Ивахненко Андрей Александрович*

shaurushka@gmail.com

Москва, Вычислительный центр им. А. А. Дородницына РАН

Для повышения обобщающей способности методов классификации необходимо оценивать частоту ошибок  $\nu(a, X)$  классификатора  $a$ , построенного по обучающей выборке  $X$ , на контрольной выборке  $\bar{X}$ . В комбинаторной теории переобучения такие оценки выводятся на основании предположения о равной вероятности всех разбиений выборки  $X \sqcup \bar{X}$  на обучающую и контрольную подвыборки. Основным объектом исследования является бинарная матрица ошибок размера  $L \times D$ , строки которой соответствуют объектам, столбцы — классификаторам. Рассматриваются матрицы ошибок с попарно различными столбцами, порождаемые параметрическими семействами классификаторов. До сих пор точные оценки удавалось получать только для модельных семейств классификаторов, непосредственно задаваемых матрицами ошибок специального вида.

В данной работе рассматриваются *цепи классификаторов*, порождаемые пороговыми решающими правилами вида  $r(x) = [f(x) > \theta]$ , где  $x$  — произвольный объект,  $f$  — вещественный признак, принимающий различные значения на объектах выборки,  $\theta$  — порог. Такие правила используются при построении конъюнктивных логических закономерностей, решающих деревьев и других логических алгоритмов классификации.

Разработан алгоритм вычисления вероятности переобучения произвольной прямой цепи, имеющий вычислительную сложность, полиномиальную по объёму выборки  $L$ , но экспоненциальную по числу участков монотонности в цепи. Предложен способ упрощения цепи путём уменьшения числа участков монотонности, при котором точность оценки ухудшается незначительно.

Полученная оценка используется для модификации критериев отбора признаков в жадных алгоритмах индукции конъюнктивных логических закономерностей. При решении вопроса о том, какой следующий признак добавить в конъюнкцию, модифицированный критерий может предпочесть признак, дающий меньший выигрыш информативности на обучающей выборке, но улучшающий оценку обобщающей способности. В экспериментах на модельных данных модифицированный критерий улучшает обобщающую способность отбираемых конъюнкций.

- [1] *Ишкина Ш. Х.* Комбинаторные оценки вероятности переобучения для некоторых связанных семейств классификаторов // Машинное обучение и анализ данных, 2013. <http://jmla.org>

## Снижение переобученности распознающих алгоритмов на основе полных решающих деревьев

*Генрихов Игорь Евгеньевич*

ingvar1485@rambler.ru

Москва, МПГУ

В отличие от классического решающего дерева (РД), в полном решающем дереве (ПРД) на каждой итерации выбирается набор признаков  $X$ , образующий так называемую полную вершину. В набор  $X$  попадают все признаки, удовлетворяющие выбранному критерию ветвления в равной или почти равной мере. Далее для каждого признака из  $X$  строится обычная внутренняя вершина по аналогии с классическим РД. В ПРД, в отличие от классического РД, описание распознаваемого объекта  $S$  может попасть в разные листья дерева. Поэтому при выборе класса, которому принадлежит объект  $S$ , используется коллективное голосование по листьям дерева. Успешное применение предложенной модели РД было ранее продемонстрировано на ряде прикладных задач.

Следует отметить, что распознающие алгоритмы на основе РД подвержены эффекту переобучения. Данная проблема переносится и на ПРД, т. к. ПРД является усложненной моделью классического РД.

В данной работе на реальных задачах исследуется эффект переобучения ПРД на примере ранее разработанного алгоритма AGI.Bias. Предложена методика снижения эффекта переобучения ПРД с использованием числовых характеристик построенного дерева, а именно, средней глубины дерева, обозначаемой  $D$ , и среднего числа обучающих объектов  $J$ , описания которых попадают в лист дерева. Первый этап процедуры — синтез начального ПРД и вычисление характеристик  $D$  и  $J$ . Второй этап процедуры — синтез итогового ПРД. Итоговое ПРД отличается от начального ПРД тем, что текущая ветвь обрывается и строится лист дерева, если глубина текущей ветви превышает значение  $D$  или число объектов в текущем множестве обучающих объектов меньше значения  $J$ . Также рассматривались варианты построения итогового ПРД с использованием одной из указанных характеристик и с небольшими отступлениями от точных значений.

Эксперименты на прикладных задачах показали эффективность предложенной методики. Наилучшие результаты показали модификации алгоритма AGI.Bias с комбинациями « $D$  и  $J$ » и « $D - 2$  и  $J + 2$ ».

Работа поддержана грантом РФФИ № 13-01-00787.

- [1] *Генрихов И. Е.* Исследование переобученности распознающих процедур на основе полных решающих деревьев // Программные продукты и системы, Тверь: НИИ ЦПС. — №4, 2011. — С. 141–147.

## Повышение эффективности алгоритма классификации на основе анализа формальных понятий

*Прокашева Ольга Владимировна*

prokasheva@gmail.com

Москва, МГУ имени М.В.Ломоносова

Анализ формальных понятий — относительно новый метод анализа данных, относящейся к прикладной ветви алгебраической теории решёток. Данный метод нашел широкое применение в различных областях машинного обучения, таких как информационный поиск, обработка документов и текстов, построение таксономий и классификации.

В данной работе исследуется подход классификации на два класса на основе ДСМ-метода генерации гипотез. В рамках данного подхода гипотезой называется некоторый набор признаков, который присутствует в описании объектов одного класса и не присутствует в описании объектов из других классов. Таким образом гипотезы способны классифицировать новые недоопределенные объекты. Согласно методу АФП гипотезы извлекаются из решёток понятий.

В настоящей работе для каждого класса строится набор всевозможных минимальных гипотез. Алгоритм классификации с использованием извлеченных из обучающей выборки гипотез тестировался на реальных данных с номинальными и вещественными признаками. Было установлено, что алгоритм в большом количестве случаев отказывается от классификации по противоречию, то есть гипотезы относят объект сразу к двум классам. Поэтому было предложено применять различные модификации алгоритма для уменьшения отказов от классификации. В частности были использованы метрики и процедура голосования по большинству. Также осуществлялся отбор гипотез, которые удовлетворяют некоторой доле  $P$  объектов из обучающей выборки.

В результате наилучшим решением для модификации простого алгоритма стало применение голосования по большинству и фильтрация гипотез. Эффективность работы алгоритма сравнивалась с алгоритмами SVM и C4.5, причем алгоритм на основе АФП показал наилучший результат. Установлено, что модифицированный алгоритм может быть успешно применен для решения задач с номинальными и бинарными признаками.

- [1] Прокашева О. В. Повышение эффективности алгоритма классификации на основе Анализа Формальных Понятий // Машинное обучение и анализ данных, 2013.

## Эволюционный метод классификация биологических объектов

*Цыганкова Ирина Александровна*

`rallada-ltd@infopro.spb.su`

Санкт-Петербург, Санкт-Петербургский институт информатики  
и автоматизации РАН

Предлагается метод классификации биологических объектов, основанный на эволюционном подходе к решению экстремальной задачи функции многих переменных [1]. Метод ориентирован на обработку многомерных массивов информации, особенностями которой являются высокая размерность признакового пространства и малый объем выборки объектов. Метод основан на предположении, что в многомерном признаковом пространстве существует некоторый базовый элемент, относительно которого формируется ранжированная последовательность объектов, разделяющая обучающую выборку на два класса. Метод реализует двойное ранжирование объектов относительно базового элемента: упорядочение объектов по классам и упорядочение объектов по возрастанию расстояния от базового элемента внутри классов. Тогда принадлежность объекта к одному из классов определяется рангом классифицируемого объекта в упорядоченном ряду объектов обучающей выборки. Поиск базового элемента проводится с использованием эволюционного подхода, реализуемого с помощью модифицированного генетического алгоритма, где в качестве исходной популяции рассматриваются сами объекты обучающей выборки. Поиск базового элемента при заданном типе его представления в многомерном пространстве признаков и выбранном варианте упорядочения классов в формируемой последовательности объектов является, по сути, самостоятельной задачей, что позволяет достаточно легко организовать параллельную реализацию алгоритма обучения.

Предлагаемый подход к классификации позволяет не проводить снижение размерности пространства признаков, что, в свою очередь, позволяет исключить потерю значимой информации и учесть внутренние связи в рассматриваемых информационных массивах. Проведенные исследования показали высокую эффективность предлагаемого метода. Погрешность классификации для тестового расчета составила 2–3%.

- [1] *Цыганкова И. А.* Эволюционный метод классификации биологических объектов // Информационные системы и технологии, Орел: ФГБОУ ВПО ОрелГТУ, №5, 2012. — С. 50–57.

## Метод анализа ансамбля многомерных разнотипных временных рядов и его применение при исследовании заболеваемости клещевым энцефалитом

*Бериков Владимир Борисович*<sup>1</sup>

*Пестунов Игорь Алексеевич*<sup>2</sup>

*Герасимов Максим Константинович*<sup>3</sup>

<sup>1</sup>berikov@math.nsc.ru, <sup>2</sup>pestunov@ict.nsc.ru, <sup>3</sup>max\_post@ngs.ru

Новосибирск, <sup>1,3</sup>Институт математики СО РАН, <sup>2</sup>Институт вычислительных технологий СО РАН

Рассматривается задача построения модели ансамбля многомерных временных рядов, где под ансамблем понимается совокупность меняющихся во времени характеристик похожих (в определенном смысле) объектов. Наблюдаемые объекты могут описываться как количественными, так и качественными (порядковыми, номинальными, булевыми) характеристиками, а интервалы наблюдений за различными объектами могут не полностью совпадать. Построенные модели позволяют агрегировать данные по всем объектам и получать надежные решения даже в случае, когда по каждому из объектов объем наблюдений сравнительно невелик.

Для решения задачи предложены алгоритмы, основанные на логических решающих функциях (деревьях решений). Разработанные алгоритмы позволяют обнаруживать логические закономерности, описывающие динамику изучаемых объектов; в отличие от известных методов анализа панельных данных, они не требуют предположений о параметрической модели распределения.

Разработанные алгоритмы были использованы при решении прикладной задачи анализа влияния природных факторов (среднемесячная температура, влажность воздуха, количество осадков, активность Солнца) на заболеваемость клещевым энцефалитом в трех эндемичных регионах (Новосибирская, Иркутская области и Горный Алтай). Выявлены общие закономерности динамики заболеваемости, определяемые схожими механизмами функционирования паразитарной системы клещевого энцефалита. При этом динамика заболеваемости отличается для различных регионов в зависимости от природных и антропогенных факторов, специфичных для данных территорий. В докладе приводятся примеры обнаруженных закономерностей.

Работа поддержана грантом РФФИ № 11-07-00346а.

- [1] Бериков В. Б., Пестунов И. А., Герасимов М. К. Анализ совокупности разнотипных временных рядов с использованием логических решающих функций // Вычислительные технологии. 2012. — Том 17, № 5. С. 12–22.

## Интеллектуальный анализ данных и знаний по стентированию коронарных артерий

*Янковская Анна Ефимовна*<sup>1</sup>

*Китлер Сергей Владимирович*<sup>2</sup>

<sup>1</sup>ayyankov@gmail.com, <sup>2</sup>svkitler@gmail.com

Томск, <sup>1</sup>Томский государственный университет, <sup>1</sup>Томский государственный архитектурно-строительный университет, <sup>2</sup>Томский государственный университет систем управления и радиоэлектроники

Для ряда слабоструктурированных областей в задачах принятия решений (ПР) используется большой объем данных и знаний. Их обработка невозможна без применения методов интеллектуального анализа данных и знаний (ИАДЗ). В настоящее время ИАДЗ уделено большое внимание ученых, среди которых особого внимания заслуживают работы научных школ Ю. И. Журавлева, Н. Г. Загоруйко, А. Д. Закревского, В. К. Финна, А. Е. Янковской, Gregory Piatetsky-Shapiro и др. Однако, несмотря на наличие ряда нижеперечисленных публикаций, задача ИАДЗ по стентированию коронарных артерий (КА) до сих пор остается открытой.

В данном исследовании излагаются основные этапы ИАДЗ, предлагается подход к анализу данных и знаний по стентированию КА, реализующая его интеллектуальная система (ИС) и результаты её исследования.

В отличие от систем «Fuzzy expert system approach for coronary artery disease screening using clinical parameters» и «A multilayer perceptron-based medical decision support system for heart disease diagnosis», основанных на мягких вычислениях и предназначенных для обнаружения и прогнозирования ишемической болезни сердца на ранней стадии и для диагностики сердечно-сосудистых заболеваний соответственно, а также системы «Knowledge Management System for Clinical Decision Support – Application in Cardiology», основанной на статистических методах и созданной для поддержки ПР по лечению больных с острым коронарным синдромом, предлагаемая ИС анализа данных и знаний по стентированию КА основана на логико-комбинаторных методах распознавания образов.

Применение ИАДЗ позволит снизить риски возникновения неблагоприятных сердечно-сосудистых событий и оптимизировать лечение больных, подвергшихся инвазивным вмешательствам на КА.

В дальнейшем предлагается осуществлять ПР с использованием отказоустойчивых диагностических тестов.

Работа частично поддержана грантами РФФИ № 13-07-00373, № 13-07-98037-р\_сибирь\_a, № 12-07-31109\_мол-а и грантом РГНФ № 13-06-00709.

- [1] Янковская А. Е., Китлер С. В. Интеллектуальный анализ данных и знаний по стентированию коронарных артерий // Машинное обучение и анализ данных, 2013 (подано в печать).

## Использование модифицированной версии метода оптимальных достоверных разбиений для анализа клинических данных

*Сенько Олег Валентинович*<sup>1</sup>

*Кузнецова Анна Викторовна*<sup>2</sup>

*Костомарова Ирина Викторовна*<sup>3</sup>

<sup>1</sup>senkoov@mail.ru, <sup>2</sup>azfor@narod.ru, <sup>3</sup>h1a2222@yandex.ru

Москва, <sup>1</sup>Вычислительный центр им. А. А. Дородницына РАН,

<sup>2</sup>ИБХФ им. Н. М. Эмануэля РАН,

<sup>3</sup>Филиал ГБОУ ВПО РНИМУ им. Н. И. Пирогова НКЦ геронтологии

На примере задачи исследования связи клинико-лабораторных показателей с развитием транзиторной ишемической атаки (ТИА) и ишемического инсульта у пациентов старших возрастных групп с хронической ишемией головного мозга рассматривается новая модификация метода интеллектуального анализа данных «Оптимальных достоверных разбиения» (ОДР). Метод ОДР предназначен для поиска по выборкам данных наборов верифицированных эмпирических закономерностей. При этом под закономерностью понимается оптимальное разбиение пространства предикторов с достоверным различием средних значений целевой переменной в соседних областях. В работе рассматривается новый способ верификации, позволяющий выделить закономерности, для которых статистически обосновывается необходимость включения для достижения максимальной обобщающей способности всех характеризующих элементов. Например, при использовании двумерных моделей обосновывается необходимость включения каждой из двух переменных. В отличие от ранее использовавшегося подхода новый способ позволяет производить верификацию для произвольных разбиений. Обсуждаются также вариант метода ОДР, направленный на поиск достоверных различий между характером зависимости целевой переменной от предикторов в различных группах объектов. Данный вариант позволил выявить достоверные различия между характером зависимости риска ишемического инсульта от биохимических показателей в группе пациентов, перенёсших ранее ТИА, и в группе без ТИА. Рассматриваются также методы анализа систем достоверных закономерностей.

Работа поддержана грантом РФФИ № 11-07-00715.

- [1] Кузнецова А. В., Костомарова И. В., Сенько О. В. Логико-статистический анализ связи клинколабораторных показателей с возникновением нарушений мозгового кровообращения у пациентов пожилого возраста с хронической ишемией головного мозга // Математическая биология и биоинформатика, 2013. — Т. 8, № 1. — С. 182–184. [http://www.matbio.org/2013/Kuznetsova\\_8\\_182.pdf](http://www.matbio.org/2013/Kuznetsova_8_182.pdf).

## Синтез простейших нелинейных систем квазиинвариантного управления с заданными свойствами методами распознавания образов

*Теклина Лариса Григорьевна*

*Котельников Игорь Вячеславович*

*Гельфер Ирина Самуиловна*

peymark@pmk.unn.ru

Нижегородский государственный университет  
им. Н. И. Лобачевского

Синтез систем квазиинвариантного управления, систем, которые должны не устранять возникающие ошибки, а предотвращать их, т. е. сделать объект управления невосприимчивым (инвариантным) к внешним воздействиям — весьма актуальная проблема. Квазиинвариантность означает малость ошибки управления в установившемся режиме. Для решения этой проблемы используется новый подход, основанный на постановке и решении задачи синтеза методами распознавания образов с активным экспериментом. Отличительной особенностью нового подхода является переход от классических методов исследования к новой методике, в которой сложные расчеты заменены достоверным математическим экспериментом с последующей обработкой результатов экспериментов методами интеллектуального анализа данных, работающими в пространствах большой размерности. На базе этого подхода разработана новая методика синтеза *линейных* систем квазиинвариантного управления. Применение методов распознавания образов позволило синтезировать системы управления с заданными свойствами, характеризующими не только установившийся режим, но и качество переходного процесса.

Однако, и теоретически, и экспериментально показано, что в рамках линейных систем удовлетворить всем требованиям, выдвигаемым к системе управления, часто бывает сложно, но возможно при выборе нелинейной стратегии управления. Работа посвящена дальнейшему развитию нового подхода к синтезу систем квазиинвариантного управления путем расширения методики синтеза линейных систем на область *нелинейных* систем управления, что позволяет преодолеть главный недостаток линейных систем — большие значения функции управления в переходном процессе.

Работа поддержана грантом РФФИ № 11-01-00379.

- [1] *Теклина Л. Г., Котельников И. В., Гельфер И. С.* Применение методов распознавания образов для синтеза кусочно-линейных систем квазиинвариантного управления // Машинное обучение и анализ данных, 2013. <http://jmla.org/papers/index.php/JMLDA/authors/submission/51>.

## О некоторых моделях «обучения» компьютера, индуцирующих NP-трудные задачи дискретной оптимизации

*Кельманов Александр Васильевич*

kelm@math.nsc.ru

Новосибирск, ИМ СО РАН

Рассматривается подход к решению проблем «обучения» компьютера распознаванию образов, состоящий в реализации следующей цепочки последовательно выполняемых шагов: 1) формулировка содержательной проблемы, 2) её формализация в виде оптимизационной модели, 3) выявление дискретной экстремальной задачи, индуцированной этой моделью, 4) построение полиномиального алгоритма решения задачи с априорно гарантированной оценкой точности, 5) компьютерная технология.

Содержательной проблеме ставится в соответствие фиксированная гипотетическая модель порождения данных. Оптимизационная модель «обучения» формулируется как задача аппроксимации данных либо по критерию минимума суммы квадратов уклонений, либо по критерию максимального правдоподобия. Аналитическими методами показывается, что оптимизация критерия сводится к решению специфической дискретной экстремальной задачи. Исследуется её вычислительная сложность. Если индуцированная задача NP-трудна, то строится полиномиальный приближенный алгоритм и доказывается его оценка точности.

Реализация подхода выявила ряд NP-трудных задач и соответствующих им проблем «обучения», которые интуитивно, но бездоказательно считались труднорешаемыми. Для этих задач построены приближенные полиномиальные алгоритмы с оценками точности. С другой стороны, этот же подход позволил установить полиномиальную разрешимость многих практически важных задач анализа данных и распознавания.

Среди труднорешаемых проблем, для которых построены алгоритмы с теоретическими гарантиями, отметим проблемы поиска непересекающихся подмножеств «похожих» (в смысле евклидовых расстояний, их квадратов, и скалярных произведений) объектов. Реализованы варианты, в которых входами задач являются как координаты векторов, так и матрицы попарных сравнений.

Работа поддержана грантами РФФИ № 12-01-00090, № 12-01-33028-мол-а-вед, № 13-07-00070, а также Целевой программой СО РАН (интеграционные проекты и 7Б и 21А).

- [1] *Кельманов А. В.* О некоторых труднорешаемых задачах кластерного анализа данных // Труды молодёжной школы-семинара «Дискретные модели и методы принятия решений», Новосибирск: ИМ СО РАН, 2013. <http://math.nsc.ru/LBRT/k4/school2013/index.html>.

## 2-приближенный алгоритм для одной задачи поиска «похожих» объектов

**Ерёмин Иван Иванович**<sup>1</sup>

*Гимади Эдуард Хайрутдинович*<sup>2</sup>

*Кельманов Александр Васильевич*<sup>3</sup>

*Пяткин Артем Валерьевич*<sup>4</sup>

*Хачай Михаил Юрьевич*<sup>5</sup>

<sup>2</sup>gimadi@math.nsc.ru, <sup>3</sup>kelm@math.nsc.ru, <sup>4</sup>artem@math.nsc.ru,

<sup>5</sup>mkhachay@imm.uran.ru

Екатеринбург, <sup>1,5</sup>Институт математики и механики им. Н. Н. Красовского  
УрО РАН; Новосибирск, <sup>2,3,4</sup>ИМ СО РАН

Рассматривается одна из простейших содержательных проблем, состоящая в выборе из множества объектов подмножества «похожих» элементов. Предполагаются заданными веса  $(a_1, \dots, a_n)$  объектов и матрица  $c = (c_{ij})$ ,  $1 \leq i, j \leq n$ , попарных сравнений объектов. Требуется найти подмножество строк матрицы, минимизирующее сумму весов выбранных объектов вместе с суммой результатов их попарных сравнений.

Проблема моделируется следующей экстремальной задачей. Даны полный неориентированный взвешенный  $n$ -вершинный граф  $G$  и натуральное число  $m < n$ . Требуется найти  $m$ -клик с минимальным (максимальным) суммарным весом входящих в неё вершин и рёбер.

В общем случае эта задача NP-трудна в сильном смысле и неаппроксимироваема [1]. Доказано, что в двух практически важных случаях задача, сохраняя труднорешаемость, обладает 2-приближёнными эффективными алгоритмами. В качестве приближённого использован точный алгоритм  $\mathcal{A}$  отыскания за время  $\mathcal{O}(n^2)$   $m$ -звезды минимального веса в графе  $G$ , модифицированном так, что веса вершин в новом графе равны нулю, а веса рёбер изменены по специальному правилу.

В случае неотрицательных весов вершин в графе  $G$  доказано, что алгоритм  $\mathcal{A}$  находит решение задачи на минимум с оценкой точности 2, которая: 1) *асимптотически достижима*, если веса рёбер удовлетворяют неравенству треугольника, 2) *достижима*, если веса рёбер — квадраты попарных расстояний между точками евклидова пространства.

2-приближённый полиномиальный алгоритм строится также для отыскания константного числа несмежных клик с заданными размерами.

Работа поддержана грантами РФФИ (№№ 12-01-00090, 12-01-00093, 12-01-33028-мол-а-вед, 13-07-00070, 13-07-00181), а также Целевой программой УрО и СО РАН (интеграционные проекты 12-01-1017/1 и 7Б).

- [1] Ерёмин И. И., Гимади Э. Х., Кельманов А. В., Пяткин А. В., Хачай М. Ю. 2-приближенный алгоритм поиска клики с минимальным весом вершин и рёбер // Труды ИММ УрО РАН, 2013, Т. 19. № 2, С. 134–143.

## FPTAS для одной труднорешаемой задачи поиска подмножества векторов

*Кельманов Александр Васильевич*<sup>1</sup>

*Романченко Семён Михайлович*<sup>2</sup>

<sup>1</sup>kelm@math.nsc.ru, <sup>2</sup>rsm@math.nsc.ru

Новосибирск, <sup>1,2</sup>ИМ СО РАН

Рассматривается следующая NP-трудная в сильном смысле (Кельманов А. В., Пяткин А. В., 2010)

**Задача VS-2** (Vector Subset 2). *Дано:* множество  $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$  векторов из  $\mathbb{R}^q$  и натуральное число  $M > 1$ . *Найти:* подмножество  $\mathcal{C} \subseteq \mathcal{Y}$  мощности  $M$  такое, что

$$\sum_{\mathbf{y} \in \mathcal{C}} \|\mathbf{y} - \bar{\mathbf{y}}(\mathcal{C})\|^2 \rightarrow \min,$$

где  $\bar{\mathbf{y}}(\mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum_{\mathbf{y} \in \mathcal{C}} \mathbf{y}$ .

Содержательная проблема («обучение» с одновременным «цензурированием») состоит в разбиении таблицы  $\mathcal{Y}$  с результатами измерения набора числовых характеристик на два таких подмножества, что одно из них —  $\mathcal{C}$  — соответствует результатам многократных измерений характеристик некоторого объекта, а другое —  $\mathcal{Y} \setminus \mathcal{C}$  — однократным результатам измерений характеристик произвольных (случайных) объектов.

Ранее (2010) авторами был предложен 2-приближённый полиномиальный алгоритм решения задачи за время  $O(qN^2)$ . Ими же построен (2012) точный псевдополиномиальный алгоритм для случая, когда координаты векторов целочисленны, а размерность  $q$  пространства фиксирована; временная сложность алгоритма есть величина  $O(qN(2MB)^q)$ , где  $B$  — максимальное абсолютное значение координаты векторов входного множества. Кроме того, для этой задачи обоснована (Шенмайер В. В., 2012) полиномиальная приближённая схема (PTAS) с временной сложностью  $O(qN^{2/\varepsilon+1}(9/\varepsilon)^{3/\varepsilon})$ , где  $\varepsilon$  — относительная погрешность алгоритма.

В настоящей работе для случая фиксированной размерности  $q$  пространства обоснована полностью полиномиальная приближённая схема (FPTAS), гарантирующая  $(1 + \varepsilon)$ -приближённое решение задачи за время  $O(qN^2(M/\varepsilon)^q)$ .

Работа поддержана грантами РФФИ № 12-01-00090, № 12-01-33028-мол-а-вед, № 13-07-00070, а также Целевой программой СО РАН (интеграционные проекты и 7Б и 21А).

- [1] *Кельманов А. В., Романченко С. М.* FPTAS для одной NP-трудной задачи поиска подмножества векторов // Дискретный анализ и исследование операций, Новосибирск: ИМ СО РАН, 2013 (в печати).

## 2-приближённый алгоритм для одной задачи поиска семейства непересекающихся подмножеств векторов

*Галашов Александр Евгеньевич*<sup>1</sup>

*Кельманов Александр Васильевич*<sup>2</sup>

<sup>1</sup>galashov.alexandr@gmail.com, <sup>2</sup>kelm@math.nsc.ru

Новосибирск, <sup>1</sup>НГУ, <sup>2</sup>ИМ СО РАН

Рассматривается следующее обобщение одной NP-трудной в сильном смысле (Кельманов А. В., Пяткин А. В., 2010) задачи.

**Задача FDVS-( $\mathbb{R}^q$ )SD** (Family of Disjoint Vector Subsets, Euclidean case for Squared Distances). *Дано:* множество  $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$  векторов из  $\mathbb{R}^q$  и натуральные числа  $M_1, \dots, M_J$ . *Найти:* семейство  $\{\mathcal{C}_1, \dots, \mathcal{C}_J\}$  непересекающихся подмножеств множества  $\mathcal{Y}$  такое, что

$$\sum_{j=1}^J \sum_{\mathbf{y} \in \mathcal{C}_j} \|\mathbf{y} - \bar{\mathbf{y}}(\mathcal{C}_j)\|^2 \rightarrow \min,$$

где  $\bar{\mathbf{y}}(\mathcal{C}_j) = \frac{1}{|\mathcal{C}_j|} \sum_{\mathbf{y} \in \mathcal{C}_j} \mathbf{y}$ , при ограничениях  $|\mathcal{C}_j| = M_j, j = 1, \dots, J$ .

Эта слабо изученная в алгоритмическом плане задача индуцируется, в частности, одной из актуальных проблем помехоустойчивого анализа данных и распознавания образов [1]. В постановочном плане задача FDVS-( $\mathbb{R}^q$ )SD близка к классической задаче MSSC, которая в публикациях нередко фигурирует под названием *k*-Means. Напомним, что содержательная проблема, индуцирующая задачу MSSC, состоит в разбиении имеющейся таблицы. При этом предполагается, что таблица не содержит результатов измерений произвольных объектов. Однако, судя по имеющимся публикациям, на практике нередки ситуации, когда табличные данные содержат «мусор» в виде случайных измерительных данных («выбросов»). Задача FDVS-( $\mathbb{R}^q$ )SD моделирует эту ситуацию. Содержательная проблема, соответствующая этой задаче, состоит в совместной кластеризации и (так называемом) цензурировании выборочных данных.

В настоящей работе обоснован 2-приближённый алгоритм, имеющий временную сложность  $\mathcal{O}(N^2(N^{J+1} + q))$ . При фиксированном  $J$  алгоритм полиномиален. Ранее лишь для простейшего, но труднорешаемого случая (когда  $J = 1$ ) этой задачи были построены алгоритмы с гарантированными оценками точности.

Работа поддержана грантами РФФИ № 12-01-00090, № 12-01-33028-мол-а-вед, № 13-07-00070, а также Целевой программой СО РАН (интеграционные проекты и 7Б и 21А).

[1] *Галашов А. Е., Кельманов А. В.* 2-приближённый алгоритм для одной задачи поиска семейства непересекающихся подмножеств векторов // Автоматика и телемеханика, Москва: Наука, 2013 (принята в печать).

## Рандомизированный алгоритм для одной NP-трудной задачи кластерного анализа

*Кельманов Александр Васильевич*<sup>1</sup>

*Хандеев Владимир Ильич*<sup>2</sup>

<sup>1</sup>kelm@math.nsc.ru, <sup>2</sup>vladimir.handeev@gmail.com

Новосибирск, <sup>1</sup>ИМ СО РАН, <sup>2</sup>НГУ

Рассматривается следующая NP-трудная в сильном смысле (Кельманов А. В., Пяткин А. В., 2009)

**Задача 1-MSSC-F.** *Дано:* множество  $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$  векторов из  $\mathbb{R}^q$  и натуральное число  $M > 1$ . *Найти:* подмножество  $\mathcal{C} \subseteq \mathcal{Y}$  мощности  $M$  такое, что

$$\sum_{\mathbf{y} \in \mathcal{C}} \|\mathbf{y} - \bar{\mathbf{y}}(\mathcal{C})\|^2 + \sum_{\mathbf{y} \in \mathcal{Y} \setminus \mathcal{C}} \|\mathbf{y}\|^2 \rightarrow \min,$$

где  $\bar{\mathbf{y}}(\mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum_{\mathbf{y} \in \mathcal{C}} \mathbf{y}$  — центр подмножества  $\mathcal{C}$ .

Задача индуцируется одной из актуальных проблем помехоустойчивого анализа данных, которая заключается в разбиении таблицы  $\mathcal{Y}$  с результатами измерения набора числовых характеристик некоторого объекта на два таких подмножества, что одно из них —  $\mathcal{C}$  — соответствует активному, а другое —  $\mathcal{Y} \setminus \mathcal{C}$  — пассивному состояниям этого объекта.

Ранее был предложен (Долгушев А. В., Кельманов А. В., 2011) 2-приближённый полиномиальный алгоритм решения задачи за время  $\mathcal{O}(qN^2)$ . Кроме того, построена полиномиальная приближённая схема (PTAS) с временной сложностью  $\mathcal{O}(qN^{2/\varepsilon+1}(9/\varepsilon)^{3/\varepsilon})$ , где  $\varepsilon$  — относительная погрешность алгоритма (Долгушев А. В., Кельманов А. В., Шенмайер В. В., 2011). В настоящей работе обоснован рандомизированный полиномиальный алгоритм, который находит  $(1 + \varepsilon_N)$ -приближённое решение с вероятностью  $1 - \delta_N$ , где  $\varepsilon_N, \delta_N \rightarrow 0$  при  $N \rightarrow \infty$ .

Суть алгоритма состоит в следующем. Из множества  $\mathcal{Y}$  случайным образом выбирается мультимножество фиксированной мощности. Для каждого из его непустых подмножеств вычисляется центр и формируется набор из  $M$  векторов исходного множества, имеющих наибольшие проекции на этот центр. Сформированный набор объявляется претендентом на решение. В качестве окончательного решения выбирается подмножество-претендент, для которого значение целевой функции минимально.

Работа поддержана грантами РФФИ № 12-01-00090, № 12-01-33028-мол-а-вед, № 13-07-00070, а также Целевой программой СО РАН (интеграционные проекты и 7Б и 21А).

- [1] *Кельманов А. В., Хандеев В. И.* Рандомизированный алгоритм для одной NP-трудной задачи кластерного анализа // Дискретный анализ и исследование операций, Новосибирск: ИМ СО РАН, 2013 (в печати).

## О сложности одной задачи кластерного анализа

*Кельманов Александр Васильевич*<sup>1</sup>

*Пяткин Артем Валерьевич*<sup>2</sup>

<sup>1</sup>kelm@math.nsc.ru, <sup>2</sup>artem@math.nsc.ru

Новосибирск, <sup>1,2</sup>ИМ СО РАН

Рассматривается следующая задача кластерного анализа.

**Задача MSSC-PD** (MSSC, the case for Pairwise Distances). *Дано:* множество  $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$  векторов из  $\mathbb{R}^q$ . *Найти:* разбиение этого множества на непустые кластеры  $\mathcal{C}_1, \dots, \mathcal{C}_J$  такое, что

$$\sum_{j=1}^J \sum_{x, z \in \mathcal{C}_j} \|x - z\|^2 \rightarrow \min.$$

На протяжении нескольких десятилетий эта задача интуитивно и гипотетически, но бездоказательно считалась NP-трудной. В настоящей работе доказано, что она NP-трудна в сильном смысле. При доказательстве показано, что к её частному случаю (когда  $J = 2$ ) полиномиально сводится следующая

**Задача Max-Cut-( $\mathbb{R}^q$ )SD** (Max-Cut in Euclidean space, for Squared Distances). *Дано:* множество  $\mathcal{Y} = \{y_1, \dots, y_N\}$  векторов из  $\mathbb{R}^q$ . *Найти:* разбиение этого множества на два подмножества  $\mathcal{X}$  и  $\mathcal{Z}$  такое, что

$$\sum_{x \in \mathcal{X}} \sum_{z \in \mathcal{Z}} \|x - z\|^2 \rightarrow \max.$$

При этом впервые доказано [1], что задача Max-Cut-( $\mathbb{R}^q$ )SD NP-трудна в сильном смысле. Идея доказательства состоит в полиномиальном сведении к частному случаю задачи Max-Cut-( $\mathbb{R}^q$ )SD классической NP-трудной в сильном смысле задачи Max-Cut.

Для целевой функции задачи MSSC-PD справедливо равенство

$$\sum_{j=1}^J \sum_{x, z \in \mathcal{C}_j} \|x - z\|^2 = 2 \sum_{j=1}^J |\mathcal{C}_j| \sum_{y \in \mathcal{C}_j} \|y - \bar{y}(\mathcal{C}_j)\|^2.$$

где  $\bar{y}(\mathcal{C}_j) = \frac{1}{|\mathcal{C}_j|} \sum_{y \in \mathcal{C}_j} y$ . Поэтому задача кластеризации по критерию минимума правой части последнего равенства NP-трудна в сильном смысле.

Работа поддержана грантами РФФИ № 12-01-00090, № 12-01-00093, № 12-01-33028-мол-а-вед, № 13-07-00070, а также Целевой программой СО РАН (интеграционные проекты и 7Б и 21А).

[1] *Кельманов А. В., Пяткин А. В.* О сложности одной задачи о разрезе максимального веса // Дискретный анализ и исследование операций, Новосибирск: ИМ СО РАН, 2013 (принята в печать).

## 2-приближённый полиномиальный алгоритм для одной NP-трудной задачи кластеризации последовательности

*Кельманов Александр Васильевич*<sup>1</sup>

*Хамидуллин Сергей Асгадуллович*<sup>2</sup>

<sup>1</sup>kelm@math.nsc.ru, <sup>2</sup>kham@math.nsc.ru

Новосибирск, <sup>1,2</sup>ИМ СО РАН

Рассматривается следующая NP-трудная в сильном смысле (Кельманов А. В., Пяткин А. В., 2013)

**Задача 1-MSSCS-F.** Дано: последовательность  $\mathcal{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)$  векторов из  $\mathbb{R}^q$ , натуральные числа  $T_{\min}$ ,  $T_{\max}$  и  $M > 1$ . Найти: подмножество  $\mathcal{M} = \{n_1, \dots, n_M\} \subseteq \mathcal{N} = \{1, \dots, N\}$  номеров элементов последовательности  $\mathcal{Y}$  такое, что

$$\sum_{j \in \mathcal{M}} \|\mathbf{y}_j - \bar{\mathbf{y}}(\mathcal{M})\|^2 + \sum_{i \in \mathcal{N} \setminus \mathcal{M}} \|\mathbf{y}_i\|^2 \rightarrow \min,$$

где  $\bar{\mathbf{y}}(\mathcal{M}) = \frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} \mathbf{y}_i$ , при ограничениях

$$1 \leq T_{\min} \leq n_m - n_{m-1} \leq T_{\max} \leq N - 1, \quad m = 2, \dots, M,$$

на элементы искомого подмножества  $\mathcal{M}$ .

Задача моделирует, в частности, одну из актуальных проблем помехоустойчивого анализа многомерных последовательностей (сигналов). Содержательная проблема заключается в разбиении таблицы с упорядоченными по времени результатами измерения набора числовых характеристик некоторого объекта на две таких подтаблицы, что одна из них соответствует активному, а другая пассивному состояниям этого объекта.

В настоящей работе обоснован приближенный алгоритм решения задачи с асимптотически достижимой оценкой точности 2. Трудоемкость алгоритма есть величина  $\mathcal{O}(N^2(M(T_{\max} - T_{\min} + 1) + q))$ , где множители  $M$  и  $(T_{\max} - T_{\min} + 1)$  не превосходят  $N$ . Поэтому алгоритм полиномиален, а его временную сложность можно оценить как  $\mathcal{O}(N^2(N^2 + q))$ .

Суть предлагаемого подхода состоит в замене решения исходной труднорешаемой задачи 1-MSSCS-F решением более простой вспомогательной задачи и последующей оценкой точности этой замены [1].

Работа поддержана грантами РФФИ № 12-01-00090, № 12-01-33028-мол-а-вед, № 13-07-00070, а также Целевой программой СО РАН (интеграционные проекты и 7Б и 21А).

[1] *Кельманов А. В., Хамидуллин С. А.* Приближённый полиномиальный алгоритм для одной NP-трудной задачи разбиения последовательности // Дискретный анализ и исследование операций, Новосибирск: ИМ СО РАН, 2013 (принята в печать).

## Задача суммирования элементов двух последовательностей

*Кельманов Александр Васильевич*<sup>1</sup>

*Михайлова Людмила Викторовна*<sup>2</sup>

<sup>1</sup>kelm@math.nsc.ru, <sup>2</sup>mikh@math.nsc.ru

Новосибирск, <sup>1,2</sup>ИМ СО РАН

Рассматривается следующая экстремальная

**Задача SES** (Sum of Elements from two Sequences). *Дано*: числовые последовательности  $g(n)$  и  $f(n)$ ,  $n \in \mathcal{N} = \{1, \dots, N\}$ , натуральные числа  $J, K, T_{\min}$  и  $T_{\max}$ . *Найти*: непересекающиеся наборы  $\mathcal{K} \subseteq \mathcal{N}$  и  $\mathcal{J} \subseteq \mathcal{N}$  номеров элементов этих последовательностей, такие, что  $|\mathcal{K}| = K$ ,  $|\mathcal{J}| = J$ ,

$$\sum_{k \in \mathcal{K}} g(k) + \sum_{j \in \mathcal{J}} f(j) \rightarrow \max,$$

а элементы объединенного набора  $\mathcal{K} \cup \mathcal{J} = \{n_1, \dots, n_{K+J}\}$  удовлетворяют ограничениям  $1 \leq T_{\min} \leq n_m - n_{m-1} \leq T_{\max} \leq N-1$ ,  $m = 2, \dots, K+J$ .

Эта задача актуальна, в частности, при решении ряда проблем механоустойчивого анализа и распознавания одномерных и многомерных временных рядов (сигналов). Вид функций  $g$  и  $f$  определяется формализацией содержательной проблемы. Задача SES возникает, например, в ситуации, когда наблюдаемый зашумленный сигнал включает повторяющийся информационно-значимый фрагмент, совпадающий с элементом из заданного алфавита векторов, а между повторами допустимы произвольные (мешающие) фрагменты-вставки, совпадающие с элементами другого заданного алфавита векторов. При этом известно, что интервал между двумя такими фрагментами ограничен сверху и снизу заданными константами. Требуется распознать последовательность как объект, включающий повторяющийся фрагмент из первого алфавита.

Ранее (Долгушев А. В., Кельманов А. В., 2009) был обоснован эффективный алгоритм, обеспечивающий точное решение задачи за время  $O(N^5)$ . В настоящей работе предложен [1] ускоренный точный полиномиальный алгоритм, реализующий схему динамического программирования и имеющий трудоемкость  $O(N^4)$ .

Работа поддержана грантами РФФИ 11-01-00696, 12-01-00090, 13-07-00070, 12-01-33028-мол-а-вед, а также грантами целевой программы СО РАН (интеграционные проекты 7Б и 21А).

- [1] *Кельманов А. В., Михайлова Л. В.* Оптимальное суммирование элементов двух последовательностей // Дискретный анализ и исследование операций, Новосибирск: ИМ СО РАН, 2013 (принята в печать).

## Оценка вычислительной сложности задачи монотонизации выборки

*Зухба Анастасия Викторовна*

a\_\_l@mail.ru

Москва, Московский физико-технический институт

Рассматривается задача классификации на два класса  $Y = \{A, B\}$ , в которой каждый объект  $x$  множества объектов  $X$  описывается набором признаков. Каждый признак индуцирует на множестве объектов  $X$  отношение линейного порядка, а любое непустое подмножество признаков индуцирует на  $X$  отношение частичного порядка. На множестве классов задан порядок  $A > B$ .

Задача состоит в построении классификатора, который был бы монотонен, т. е. по заданному множеству признаков присваивал бы объектам ответы так, что если  $x \leq x'$ , то и ответ на  $x$  не больше чем на  $x'$ .

В качестве подготовки данных для построения монотонного классификатора рассматривается задача *монотонизации* множества. Задача состоит в отборе объектов и признаков так, чтобы для любых  $a \in A$  и  $b \in B$  объект  $a$  был бы больше либо несравним с  $b$  по совокупности признаков. Пары объектов  $a_i \in A$ ,  $b_i \in B$  такие, что  $a_i > b_i$ , будем называть *монотонными*; пары  $a_i < b_i$  будем называть *дефектными*.

Вводятся следующие характеристики качества монотонизации: количество монотонных пар, количество дефектных пар, количество объектов, количество признаков. Задачи отбора признаков и отбора объектов рассматриваются отдельно. Если брать различные сочетания этих характеристик в качестве критерия оптимальности и ограничений типа неравенств, то получается серия постановок задач дискретной оптимизации. В докладе рассматриваются вопросы их вычислительной сложности.

Между критериями оптимальности устанавливается связь, которая позволяет разделить эти задачи на подгруппы. При этом сразу выделяются тривиальные задачи, для которых строятся полиномиальные алгоритмы. Построение полиномиального алгоритма для одной из постановок основано на известной задаче о паросочетании. Большинство задач монотонизации являются NP-полными. Доказательство NP-полноты основано на сведении к ним известных NP-полных задач о рюкзаке и о покрытии множества подмножествами.

- [1] *Зухба А. В.* Оценка вычислительной сложности задачи монотонизации выборки // Машинное обучение и анализ данных, 2013. <http://jmla.org>

## Применение алгоритма поиска с запретами в задаче автоматизированного составления оптимального штатного расписания

*Боровых Наталья Игоревна*<sup>1</sup>

*Красоткина Ольга Вячеславовна*<sup>2</sup>

<sup>1</sup>natali.bor1992@mail.ru, <sup>2</sup>o.v.krasotkina@yandex.ru

Тула, Тульский государственный университет

В статье рассматривается актуальная задача составления оптимального штатного расписания на примере составления расписания работы кассиров в магазине самообслуживания. Выполнена математическая постановка задачи как задачи линейного целочисленного программирования с ограничением на бинарность целевых переменных. Рассмотрены существующие методы решения задачи составления расписания: точные и приближенные. Выбран приближенный метаэвристический метод — поиск с запретами. Разработано приложение на ЯП Java, реализующее выбранный метод с применением библиотеки OptaPlanner, и произведено его экспериментальное исследование: сравнение близости полученного результата к решению на основе точного метода ветвей и границ, составление и анализ недельного расписания.

Полученные результаты могут быть использованы на практике для составления штатного расписания в организациях массового обслуживания, что позволит организациям уменьшить затраты по привлечению сотрудников, и тем самым увеличить свою экономическую эффективность.

- [1] *Боровых Н. И., Красоткина О. В.* Применение алгоритма поиска с запретами в задаче автоматизированного составления оптимального штатного расписания // Известия Тульского государственного университета. Технические науки, Тула: ТулГУ, 2013. — принято в печать.

## Об одной вариации метода экспоненциального взвешивания для решения задач стохастической онлайн оптимизации

*Гасников Александр Владимирович*<sup>1</sup>

*Бондаренко Александр Сергеевич*<sup>2</sup>

<sup>1</sup>gasnikov@yandex.ru, <sup>2</sup>alexander.s.bondarenko@gmail.com

Москва, ПреМоЛаб МФТИ

Рассмотрим задачу стохастической онлайн оптимизации

$$\frac{1}{N} \sum_{k=1}^N \mathbb{E}_{\xi} [f_k(x; \xi^k)] \rightarrow \min_{x \in S_n(1)}, \quad S_n(1) = \left\{ x \geq 0: \sum_{i=1}^n x_i = 1 \right\}. \quad (1)$$

К такому виду приводятся, в частности, задачи о многоруких бандитах, взвешивании экспертных решений, статистическом агрегировании оценок (Cesa-Bianchi, Lugoshi, Audibert, Bubeck, Юдицкий–Назин–Цыбаков–Ваятис). Обобщение, которое предлагается в данной работе, заключается в том, что не предполагается независимость с. в.  $\xi^k$ .

Будем считать, что  $\mathbb{E}_{\xi} [f_k(x; \xi^k)]$  — выпуклые по  $x$  функции; существует такой вектор (субградиент)  $\nabla_x f_k(x; \xi^k)$ , что

$$\|\nabla_x f_k(x; \xi^k)\|_{\infty} \leq M,$$

$$\mathbb{E} (\nabla_x f_k(x; \xi^k) - \nabla_x \mathbb{E}_{\xi} [f_k(x; \xi^k)]) \mid \Xi^{k-1} \equiv 0,$$

где  $\Xi^{k-1}$  —  $\sigma$ -алгебра, порожденная случайными величинами  $\xi^1, \dots, \xi^{k-1}$ .

Для решения задачи (1) воспользуемся адаптивным методом зеркального спуска в форме Ю. Е. Нестерова, предложенного не для онлайн случая. Положим  $x_i^1 = 1/n$ ,  $i = 1, \dots, n$ . Пусть  $t = 1, \dots, N - 1$ .

В работе показано, что для метода двойственных усреднений

$$x_i^{t+1} = \frac{\exp\left(-\frac{1}{\beta_{t+1}} \sum_{k=1}^t \frac{\partial f_k(x^k; \xi^k)}{\partial x_i}\right)}{\sum_{l=1}^n \exp\left(-\frac{1}{\beta_{t+1}} \sum_{k=1}^t \frac{\partial f_k(x^k; \xi^k)}{\partial x_l}\right)}, \quad i = 1, \dots, n, \quad \beta_t = \frac{M\sqrt{t}}{\sqrt{\ln n}}$$

справедливо следующее основное утверждение:

$$\frac{1}{N} \sum_{k=1}^N \mathbb{E} [f_k(x^k; \xi^k)] - \min_{x \in S_n(1)} \frac{1}{N} \sum_{k=1}^N \mathbb{E}_{\xi} [f_k(x; \xi^k)] \leq 2M \sqrt{\frac{\ln n}{N}}.$$

Работа поддержана грантами РФФИ, Президента РФ и ПреМоЛаб.

- [1] *Гасников А. В., Нестеров Ю. Е., Спокойный В. Г.* Об эффективности одного метода рандомизации зеркального спуска в задачах online оптимизации // Автоматика и телемеханика. 2014. (подана)

## Об одном методе отсечений с отбрасыванием отсекающих плоскостей

*Заботин Игорь Ярославич*<sup>1</sup>

*Яруллин Рашид Саматович*<sup>2</sup>

<sup>1</sup>IYaZabotin@mail.ru, <sup>2</sup>YarullinRS@gmail.com

Казань, Казанский (Приволжский) федеральный университет

Предлагается метод отсечений для решения задачи математического программирования [1]. Метод характерен тем, что не требует вложения каждого из погружающих множеств в предыдущее. Такая особенность дает определенные преимущества методу перед некоторыми известными алгоритмами отсечений при его практической реализации.

Пусть  $f_j(x)$ ,  $j \in J = \{1, \dots, m\}$  — выпуклые в  $\mathbb{R}^n$  функции, множества  $D_j = \{x \in \mathbb{R}^n : f_j(x) \leq 0\}$  таковы, что  $\text{int } D_j \neq \emptyset \forall j \in J$ ,  $D' = \bigcap_{j \in J} D_j$ ,  $D'' \subset \mathbb{R}^n$  — выпуклое замкнутое множество,  $f(x)$  — непрерывная функция, достигающая на  $D = D' \cap D''$  минимального значения  $f^*$ . Решается задача  $\min\{f(x) : x \in D\}$ . Положим  $X^* = \{x \in D : f(x) = f^*\}$ ,  $F(x) = \max\{f_j(x), j \in J\}$ ,  $D'_\varepsilon = \{x \in \mathbb{R}^n : F(x) \leq \varepsilon\}$ ,  $\varepsilon \geq 0$ .

Предлагаемый метод заключается в следующем. Строится выпуклое замкнутое множество  $M_0 \subset \mathbb{R}^n$ , содержащее хотя бы одну точку множества  $X^*$ , например,  $x^*$ . Выбираются точки  $v^j \in \text{int } D_j$  для всех  $j \in J$ . Задается  $\varepsilon_0 > 0$ . Полагается  $k = 0$ ,  $i = 0$ .

1. Находится точка  $y_i \in M_i \cap D'' \cap \{x \in \mathbb{R}^n : f(x) \leq f^*\}$ .
2. Формируется  $J_i = \{j \in J : y_i \notin D_j\}$ . Если  $J_i = \emptyset$ , то  $y_i \in X^*$ .
3. Если  $y_i \notin D'_{\varepsilon_k}$ , то  $Q_i = M_i$ , и следует переход к п. 4. В противном случае полагается  $i_k = i$ ,  $x_k = y_{i_k}$ , и выбирается выпуклое замкнутое множество  $Q_i = Q_{i_k}$  такое, что  $x^* \in Q_i$ . Задается  $\varepsilon_{k+1} > 0$ , значение  $k$  увеличивается на единицу, и следует переход к очередному пункту.
4. Для каждого  $j \in J_i$  в интервале  $(v^j, y_i)$  выбирается  $z_i^j \notin \text{int } D_j$  так, чтобы при некотором  $q_i^j \in [1, q]$ ,  $q < +\infty$ , для точки  $\bar{y}_i^j = y_i + q_i^j(z_i^j - y_i)$  выполнялось  $\bar{y}_i^j \in D_j$ . Для всех  $j \in J \setminus J_i$  полагается  $z_i^j = \bar{y}_i^j = y_i$ .
5. Для каждого  $j \in J_i$  выбирается множество  $A_i^j$  обобщенно-опорных векторов к  $D_j$  в точке  $z_i^j$ , полагается  $M_{i+1} = Q_i \cap \{x \in \mathbb{R}^n : \langle a, x - z_i^j \rangle \leq 0$

$\forall a \in A_i^j\}$ , и следует переход к п. 1 при  $i$ , увеличенном на единицу.

Доказано, что при  $\varepsilon_k \rightarrow 0$ ,  $k \rightarrow \infty$ , любая предельная точка последовательности  $\{x_k\}$  принадлежит  $X^*$ .

- [1] *Заботин И. Я., Яруллин Р. С.* Об одном подходе к построению алгоритмов отсечений с отбрасыванием отсекающих плоскостей // Изв. вузов. Матем. — 2013. — № 3. — С. 74–79.

## Иерархические и оптимальные приближения цифрового изображения

*Харинов Михаил Вячеславович*

*khar@iias.spb.su*

Санкт-Петербург, СПИИРАН

В докладе развивается подход [1] к сегментации цифровых изображений посредством минимизации суммарной квадратичной ошибки  $E$  или среднеквадратичного отклонения изображения от его кусочно-постоянного приближения. Основой подхода является расширенная модель Мамфорда–Шаха, в которой утраивается число операций со связными сегментами и, помимо сегментов, рассматриваются приближения изображения кластерами из несмежных сегментов, а также кластерами из не обязательно связных пикселей изображения. Основным нововведением в модели Мамфорда–Шаха является аналитически обоснованный метод кластеризации, в частности, сегментации, который по рассматриваемой стандартной оценке качества разбиений превосходит традиционный метод  $K$ -средних и применим в качестве многомерного аналога метода Оцу. Характерный для нашего метода алгоритм кластеризации выводится из условий минимизации  $E$  и характеризуется как алгоритм наращивания кластеров, в котором именно соображениями минимизации целевого функционала определяется, как порождаются кластеры, как они разделяются на составные части и как обмениваются составными частями друг с другом. Результатом вычислений являются приближения изображения, для которых значение суммарной квадратичной ошибки  $E$  монотонно спадает от максимального (при единственном кластере) до нуля при достаточном числе кластеров. Разбиения изображения, вычисленные для последовательного числа кластеров, не зависят от линейного преобразования яркости, а алгоритмы генерации приближений коммутируют с увеличением размеров изображения дублированием пикселей.

Очередной результат состоит в аппроксимации последовательности перекрывающихся оптимальных приближений иерархической последовательностью приближений. При этом отличие приближений от изображения в зависимости от числа кластеров описывается выпуклой последовательностью значений  $E$ , что способствует их минимизации.

Оптимизация качества последовательных приближений, сохраняющаяся при масштабировании изображения по яркости и координатам, перспективна для приложений в задачах распознавания и устойчивой передачи информации.

- [1] *Харинов М. В.* Обобщение трех подходов к оптимальной сегментации цифрового изображения // Труды СПИИРАН, Санкт-Петербург: Наука, 2013., № 2(25). — С. 294–316.

## Алгоритм цифровой локализации изображений символьных меток на основе одномерного дифференциально-интегрального исчисления яркости

*Орлов Алексей Александрович*<sup>1</sup>

*Астафьев Александр Владимирович*<sup>2</sup>

<sup>1</sup>alexeyalexorlov@gmail.com, <sup>2</sup>Alexandr.astafiev@isconnection.ru

Муром, МИ (филиал) ВлГУ

С активным развитием систем технического зрения (СТЗ) для целей автоматизации производственных процессов появляются проблемы выбора и создания алгоритмов для поиска, локализации и идентификации промышленных объектов.

Целью данного исследования является разработка и внедрение алгоритма локализации символьных меток на основе одномерного дифференциально-интегрального исчисления яркости на производственных объектах металлообрабатывающей отрасли.

В сфере металлообработки зачастую используется методика маркировки изделий с использованием металлических пластин со штрих-кодом типа (Штрих-код Code 39), которые закрепляются на объектах производства — слябах.

Согласно заданного типа маркировки необходимо спроектировать алгоритм локализации её на изображении. Исходя из того, что большая часть маркера состоит из белого цвета, а остальная часть — чёрного. Это отличительный признак характеризуется сильными скачками яркости от черного к белому и от белого к черному.

Если сделать горизонтальное сечение изображения и построить график зависимости амплитуды яркости вдоль горизонтального сечения изображения, область пластины маркера на графике отразится характерными скачками. Данное свойство послужило основой для разработки алгоритма.

Разработанный алгоритм позволяет с высокой скоростью определить предполагаемые области расположения пластины маркера. Кроме того, необязательно сканировать каждую строчку исходного изображения. Можно задать минимально возможную ширину пластины и проверять сечения сделанные с интервалом в половину минимальной ширины.

Работа поддержана грантом РФФИ № 13-07-00845а.

- [1] *Астафьев А. В., Орлов А. А.* Реализация и применение алгоритма цифровой локализации изображений символьных меток на основе анализа скорости изменения яркости // Современные проблемы науки и образования, №6. — 2012. — <http://www.science-education.ru/106-7860>.

## Адаптивное модифицированное обобщённое трехмерное преобразование Хафа

*Роженцов Алексей Аркадьевич*<sup>1</sup>

*Морозовский Кирилл Валерьевич*<sup>2</sup>

<sup>1</sup>RozhencovAA@volgatchechnet, <sup>2</sup>daimond025@yandex.ru

Йошкар-Ола, Поволжский государственный технологический университет

Предлагается подход к обработке трехмерных изображений на базе модифицированного обобщённого преобразования Хафа.

Показана возможность обработки изображений с неизвестными параметрами смещения, вращения и масштабирования, а также изображений, представленных отдельными фрагментами исходного образа. Анализ эффективности алгоритма показал, что он обеспечивает вероятность правильного распознавания близкую к характеристикам пространственно-корреляционных алгоритмов.

Модифицированное обобщённое преобразование Хафа базируется на использовании трехмерного аккумуляторного массива и позволяет определять координаты центра формы, независимо от параметров вращения и масштабирования. Для уточнения положения центра формы в аккумуляторном массиве предложен адаптивный двухэтапный алгоритм преобразования Хафа. На первом этапе производится грубая оценка положения центра формы, на втором и последующем этапах область поиска постепенно сужается и поиск центра формы производится с более мелким шагом дискретизации в окрестности ранее найденного пика аккумуляторного массива. Уточнение положения центра формы прекращается после достижения допустимой погрешности оценки его координат в аккумуляторном массиве. Экспериментально показано, что предложенный алгоритм является сходящимся.

- [1] *Роженцов А. А. Морозовский К. К. Баев; А. А.* Модифицированное обобщённое преобразование Хафа для обработки трёхмерных изображений с неизвестными параметрами вращения и масштабирования // Автометрия, Новосибирск: Изд-во Сибирского отделения РАН, 2013. №2 — С. 30–41. <http://www.sibran.ru/journals/issue.php?ID=149649>.

## Цифровые изображения на комплексном дискретном торе

*Мнухин Валерий Борисович*

`mnukhin.valeriy@mail.ru`

Таганрог, Южный Федеральный Университет

В настоящее время разработка методов решения задач обработки изображений проводится, как правило, в предположении непрерывности заданного изображения. Это позволяет использовать для решения поставленных задач мощный аппарат математического анализа и интегральных преобразований. Однако на практике применение полученных таким образом решений приводит к систематическим ошибкам, связанным с дискретностью реальных цифровых изображений.

В качестве примера подобной ситуации укажем на хорошо известный метод преобразования Фурье–Меллина, используемый, в частности, при решении задач регистрации изображений. Метод основан на рассмотрении Фурье–спектра исходного изображения в полярно-логарифмической системе координат, после чего к преобразованному таким образом спектру вторично применяется преобразование Фурье. Нетрудно показать, что в *непрерывном случае* это позволяет легко определять характеристики изображений, инвариантные относительно сдвигов, вращений и масштабирования. Вместе с тем в *дискретном* случае использование этого метода осложнено невозможностью адекватного преобразования цифровых изображений в полярно-логарифмическую систему координат.

В связи с этим возникает задача разработки методов, изначально ориентированных на цифровые изображения. В данной работе цифровые изображения размера  $p \times p$ , где  $p \geq 3$  — простое число вида  $p = 4k - 1$ , предлагается рассматривать как функции на комплексных дискретных торах, которые можно считать «конечными комплексными плоскостями». Вводится понятие комплексного вращения, и строятся дискретные аналоги преобразований Меллина и Фурье–Меллина, позволяющие получать инварианты цифровых изображений относительно циклических сдвигов, масштабирования и комплексных вращений. Сохранение ортогональности при комплексных вращениях позволяет использовать их, в частности, в задачах анализа симметрии цифровых изображений.

Работа поддержана грантами РФФИ № 11-07-00591 и № 13-07-00327.

- [1] *Мнухин В. Б.* Цифровые изображения на комплексном дискретном торе // Машинное обучение и анализ данных, 2013. — С. 1–10. <http://jmla.org/papers/doc/2013/no6/Author2013Title.pdf>.

## Анализ трехмерных текстур с позиции стохастической геометрии и функционального анализа

*Федотов Николай Гаврилович*

*Голдуева Дарья Алексеевна*

nikolayfedotov@mail.ru, daria-a-m@yandex.ru

Пенза, Пензенский государственный университет

Большинство методов распознавания трехмерных объектов, как правило, предполагают предварительное упрощение анализируемого трехмерного объекта, что ведет к потере существенной доли полезной информации и, как следствие, к снижению точности анализа 3D изображений.

В настоящей работе предлагается новый универсальный подход к анализу трехмерных текстур, основанный на аппарате стохастической геометрии и функционального анализа, позволяющий анализировать трехмерные текстуры непосредственно, без предварительного их упрощения.

Кроме того, большинство методов анализа трехмерных объектов оперируют небольшим количеством признаков, имеющих конкретную интерпретацию в терминах решаемой задачи. Предлагаемый в настоящей статье метод анализа трехмерных текстур позволяет в режиме автоматической генерации формировать десятки тысяч признаков изображений, что повышает надежность их классификации.

Зачастую анализируемые трехмерные текстуры, относящиеся к одному классу, имеют высокий уровень вариабельности формы и положения повторяющихся включений. Поэтому для их распознавания целесообразно применять признаки, инвариантные к группам движений и масштабным преобразованиям.

В [1] приведены примеры триплетных признаков, инвариантных к группе движения и масштабным преобразованиям трехмерных текстур, и экспериментально определена степень их нестабильности на трехмерных текстурах, полученных с помощью атомно-силового микроскопа. Проведенные эксперименты показали, что построенные триплетные признаки обладают достаточно высокой степенью инвариантности к группе движений и к масштабным преобразованиям трехмерных текстур.

Работа поддержана грантом РФФИ № 12 07 00501.

- [1] *Федотов Н. Г., Голдуева Д. А.* Анализ трехмерных текстур с позиции стохастической геометрии и функционального анализа // Машинное обучение и анализ данных. <http://jmla.org/papers/index.php/JMLDA/author/submission/46>.

## Использование характера распределения самоподобия в качестве признака цифрового изображения в задаче классификации

*Жизняков Аркадий Львович*<sup>1</sup>

*Привезенцев Денис Геннадьевич*<sup>2</sup>

<sup>1</sup>lvovich@newmail.ru, <sup>2</sup>dgprivezencev@mail.ru

Муром, <sup>1</sup>МИ (филиал) ВлГУ, <sup>2</sup>МИ (филиал) ВлГУ

В настоящее время одним из развивающихся и перспективных направлений в цифровой обработке изображений является применение фрактального анализа. Для создания новых методов фрактальной обработки изображений необходима модель изображений. Авторами разработана фрактальная модель цифрового изображения, основанная на системах итерируемых функций.

Для качественной классификации изображений необходим набор признаков, однозначно характеризующих изображение. Чем обширнее этот набор, тем выше степень точности классификации. Поэтому кроме традиционных признаков, использование фрактальных характеристик изображений позволит снизить ошибки и существенно увеличить эффективность распознавания.

Предлагаются новые признаки изображений, характеризующие внутреннее распределение самоподобия и наиболее подобные участки изображения. Распределение самоподобия дает возможность использовать в качестве информативных признаков изображения наиболее характерные его участки, полученные на основе гистограммы характера распределения, которые в большинстве случаев являются уникальными для каждого изображения или класса изображений, и коэффициенты уравнения, описывающего характер распределения локальных признаков самоподобия.

В результате исследования было определено, что распределение самоподобия изображения инвариантно к различным преобразованиям. Коэффициенты уравнения, описывающего распределение самоподобия близки для изображений одного класса, но различны для изображений разных классов.

Работа поддержана грантом РФФИ № 13-07-00825.

- [1] *Жизняков А. Л., Привезенцев Д. Г.* Использование характера распределения самоподобия в качестве признака цифрового изображения в задаче классификации // Цифровая обработка сигналов, Москва: Российское научно-техническое общество радиотехники, электроники и связи им. А.С. Попова, №3. — 2012. — С. 64–67. [http://www.dsps.ru/abstracts/ab-312.php#an312\\_12](http://www.dsps.ru/abstracts/ab-312.php#an312_12).

## Оптимизация ядра диффузии ошибки для растривания полутоновых изображений

*Федосеев Виктор Андреевич*

vicanfed@gmail.com

Самара, Институт систем обработки изображений РАН, Самарский государственный аэрокосмический университет имени академика С. П. Королёва (национальный исследовательский университет)

Предметом исследования в данной работе является алгоритм диффузии ошибки, используемый для растривания полутоновых изображений. Работа посвящена поиску весовых функций (являющихся параметром данного алгоритма), формирующих наилучшие результаты при различных ограничениях на количество ненулевых коэффициентов и их множество значений. В качестве объективной меры качества использовался показатель WSNR, выбранный на основе результатов сравнения с другими мерами. Задача многомерной оптимизации решалась численно с использованием нескольких известных алгоритмов: Нелдера–Мида, BFGS и др.

В результате исследования наилучший результат показала следующая весовая функция, позволяющая обеспечить выигрыш по качеству около 5% в сравнении с известной функцией Флойда:

$$\begin{pmatrix} 0 & 0 & \odot & 0,5423 & 0,0533 \\ 0,0246 & 0,2191 & 0,4715 & -0,0023 & -0,1241 \\ -0,0065 & -0,0692 & 0,0168 & -0,0952 & -0,0304 \end{pmatrix}.$$

Также найдены и другие функции, позволяющие снизить вычислительную сложность алгоритма без снижения качества результата.

Поскольку метод диффузии ошибки, существенно превосходя по качеству получаемых образцов используемые в полиграфии методы, всё же не находит широкое применение в этой области из-за относительно низкой скорости работы, полученные в работе результаты могут способствовать более активному практическому применению данного метода.

Работа выполнена при частичной финансовой поддержке грантов РФФИ (проекты 12-01-00822, 13-01-97007, 12-07-31056, 13-01-12080, 13-07-12103), программы фундаментальных исследований Президиума РАН «Фундаментальные проблемы информатики и информационных технологий» (проект 2.12) и министерства образования и науки Российской Федерации (в соответствии с постановлением Правительства России от 09.04.2010 № 218).

- [1] *Федосеев В. А.* Оптимизация ядра диффузии ошибки для растривания полутоновых изображений // Компьютерная оптика, Самара, 2013. — Т. 37, № 3. — (принята к публикации)

## Формирование признакового описания фактуры картин

*Мурашов Дмитрий Михайлович*<sup>1</sup>

*Березин Алексей Владимирович*<sup>2</sup>

*Иванова Екатерина Юрьевна*<sup>3</sup>

<sup>1</sup>d\_murashov@mail.ru, <sup>2</sup>berezin\_alex@mail.ru

Москва, <sup>1</sup>Вычислительный центр им. А. А. Дородницына РАН,

<sup>2,3</sup>Государственный Исторический музей

Рассматривается задача формирования признакового пространства для сравнения изображений в атрибуции произведений живописи. Важнейшим техническим признаком картины является фактура, определяемая как видимое поверхностное строение картины и характеризующая манеру живописи художника. В соответствии с рекомендациями искусствоведов предлагается в качестве образцов для сравнения использовать группу мазков, формирующих какую-либо деталь картины. Исходными данными являются цветные изображения фрагментов картин (портретов), выполненных разными художниками 18–19 веков.

Для описания индивидуальных особенностей техники живописи предлагается использовать следующие признаки: локальная ориентация хребтов полутонного рельефа изображения, локальные признаки на основе структурного тензора (локальная ориентация градиента яркости и когерентность); локальные волновые числа. Локальная ориентация и когерентность характеризуют манеру живописи художника, специфическую для конкретных деталей картины. Локальное волновое число связано с геометрическими характеристиками кисти, которую использовал автор в процессе формирования красочного слоя рассматриваемого фрагмента (толщина волоса, ширина кисти). Для сравнения фактуры фрагментов изображений применяются методы статистического анализа.

Получены результаты тестирования признаков фактуры однотипных фрагментов двух портретов кисти Ф.Рокотова и двух портретов других авторов. Результаты показали, что предложенное признаковое описание может быть использовано для сравнения фактуры в качестве одного из видов технико-технологических исследований картин. Дальнейшие исследования будут направлены на расширение признакового описания фактуры, увеличение базы изображений картин, накопление экспериментальных данных и построение алгоритма формирования решения о схожести манеры живописи художников.

Работа поддержана грантом РФФИ № 12-07-00668.

- [1] *Мурашов Д. М., Березин А. В., Иванова Е. Ю.* Формирование признакового описания фактуры картин // Машинное обучение и анализ данных, 2013. <http://jmla.org> (представлена в редакцию).

## Учёт дальнедействующих зависимостей в задаче семантической сегментации трёхмерных облаков точек с помощью последовательной классификации

*Шаповалов Роман Викторович*<sup>1</sup>

*Ветров Дмитрий Петрович*<sup>2</sup>

*Коли Пушмит*<sup>3</sup>

<sup>1</sup>shapovalov@graphics.cs.msu.su, <sup>2</sup>vetrovd@yandex.ru,

<sup>3</sup>pkohli@microsoft.com

<sup>1,2</sup>Московский государственный университет им. М.В. Ломоносова,

<sup>3</sup>Microsoft Research Cambridge

В статье рассматривается подход к решению задачи семантической сегментации трёхмерных облаков точек, учитывающий семантический контекст, то есть зависимость меток точек от меток других точек. Типичным инструментом для учёта семантического контекста являются ненаправленные графические модели, такие как Марковские сети, однако они позволяют учитывать только локальные зависимости, иначе вывод оптимальной разметки становится вычислительно неподъёмным. В данной статье предлагается использовать аппарат последовательной классификации для учёта дальнедействующих зависимостей в трёхмерных данных.

Методы последовательной классификации итеративно классифицируют локальные признаки точек, на каждой итерации добавляя к вектору признаков метки других точек, полученные на предыдущей итерации. Особенностью предлагаемого метода является факторизация обучаемых зависимостей по пространственному признаку на так называемые *d-факторы*, которые могут вносить различный вклад в решающее правило. Настройка коэффициентов, регулирующих этот вклад, осуществляется на обучающей выборке с применением  $L_1$  регуляризации, что позволяет обнулить значительную часть коэффициентов. Таким образом, метод способен отбирать существенные зависимости в данных, игнорируя шумовые.

Эксперименты на сканах, полученных сенсором Kinect, показывают, что учёт дальнедействующих зависимостей позволяет улучшить качество сегментации. Точность предложенного метода сопоставима с точностью передовых методов сегментации облаков точек, при этом вывод работает значительно быстрее.

Работа поддержана грантами РФФИ № 12-01-00938, 12-01-33085 и грантом СП-204.2012.5.

- [1] *Shapovalov R., Vetrov D., Kohli P.* Spatial Inference Machines // Conference on Computer Vision and Pattern Recognition, Portland, OR: IEEE, 2013. <http://shapovalov.ro/papers/SIM-Shapovalov-et-al-CVPR2013.pdf>.

## Иерархическое марковское случайное поле для задачи сегментации изображений

*Осокин Антон Александрович*<sup>1</sup>

*Коли Пушмит*<sup>2</sup>

*Йегелька Штефани*<sup>3</sup>

<sup>1</sup>osokin@bayesgroup.ru, <sup>2</sup>pkohli@microsoft.com,

<sup>3</sup>stefje@eecs.berkeley.edu

<sup>1</sup>МГУ им. М.В.Ломоносова, Москва, <sup>2</sup>Майкрософт Ресерч, Кэмбридж,

<sup>3</sup>Калифорнийский университет, Беркли

Одним из наиболее успешных подходов к решению задач сегментации изображений является подход, основанный на минимизации функции дискретных переменных — энергии. Чаще всего используются парно-сепарабельные субмодулярные энергии, которые могут быть точно и эффективно минимизированы при помощи алгоритмов разрезов графов. Можно показать, что минимизация таких энергий соответствует минимизации длины границы между объектом и фоном. Данный эффект не позволяет выделять объекты, обладающие сложной границей и длинными тонкими частями.

В 2011 году Jegelka и Vilmes предложили модель «кооперативных разрезов», позволяющую частично решить данную проблему. Модель «кооперативных разрезов» является марковским полем с потенциалами высоких порядков, которые штрафуют не длину границы, а количество типов границы. Авторы показали, что задача минимизации энергии данной модели является NP-трудной, и предложили приближенный алгоритм ее решения.

В настоящей работе модель «кооперативных разрезов» переформулирована в виде иерархического марковского случайного поля, энергия которого при фиксации небольшого числа переменных становится парно-сепарабельной и субмодулярной. На основе данной формулировки построен точный неполиномиальный алгоритм минимизации энергии, позволяющий достаточно быстро решать задачу во многих практически важных случаях, а также приближенные жадные алгоритмы, превосходящие алгоритм Jegelka и Vilmes по качеству получаемого решения.

Построенные алгоритмы применены для задач бинарной и семантической сегментации, проведено экспериментальное и теоретическое исследование построенных алгоритмов.

Работа поддержана грантом РФФИ № 12-01-31254.

- [1] *Pushmeet Kohli, Anton Osokin, Stefanie Jegelka* A Principled Deep Random Field Model for Image Segmentation // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, 2013. [www.pamitc.org/cvpr13](http://www.pamitc.org/cvpr13).

## Обучение модели формы для сегментации изображений по слабо размеченным данным

*Янгель Борис Константинович*<sup>1</sup>

*Ветров Дмитрий Петрович*<sup>2</sup>

<sup>1</sup>boris.jangel@gmail.com, <sup>2</sup>vetrovd@yandex.ru

Московский Государственный Университет им. М. В. Ломоносова

В данной работе рассматривается проблема учета предпочтений на форму объекта в задаче бинарной сегментации изображений. Для ее решения предложено расширение классической модели на основе условных случайных полей моделью формы, описывающей объект графом специального вида. Для расширенной модели приведено два алгоритма нахождения оптимальной сегментации: приближенный алгоритм, основанный на стохастическом поиске, и точный, но более медленный алгоритм на основе метода ветвей и границ. Эксперименты на синтетических и реальных изображениях подтверждают, что предлагаемая модель за счет учета ограничений на форму позволяет существенно повысить качество сегментации.

Помимо этого в работе предложена процедура обучения параметров модели по слабо размеченной выборке, в которой для изображений задана только конфигурация графа формы. Метки отдельных пикселей изображений выборки при этом не известны и выступают в роли скрытых переменных. Эксперименты подтверждают, что в ходе процедуры обучения значения скрытых переменных сходятся к разумной разметке изображения, так что получаемая в результате модель фактически обучается по сильной разметке.

- [1] *B. Yangel, D. Vetrov* Learning a Model for Shape-Constrained Image Segmentation from Weakly Labeled Data // 9th Intl. Conf. on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR 2013), Springer-Verlag Berlin Heidelberg, 2013. — pp. 137–150.

## Адаптивный параметрический алгоритм обработки изображений

*Грачева Инесса Александровна*<sup>1</sup>

*Копылов Андрей Валериевич*<sup>2</sup>

<sup>1</sup>gia1509@mail.ru, <sup>2</sup>And.Kopylov@gmail.com

Тула, Тульский государственный университет

Получивший широкое распространение в анализе данных байесовский принцип оценивания приводит к единой постановке целого класса задач обработки изображений как задач поиска максимума апостериорной плотности распределения скрытого поля при фиксированных наблюдениях, причем в подавляющем большинстве случаев предполагается, что подлежащие оцениванию элементы поля принимают значения из конечного множества. Такое предположение помещает задачи такого типа в класс задач разметки, для решения которых существуют алгоритмы, имеющие полиномиальную вычислительную сложность. В данных алгоритмах так или иначе присутствует перебор по всем возможным значениям каждого оцениваемого элемента, что приводит к повышенным требованиям к объему памяти и существенно замедляет обработку, при большом количестве возможных значений, обусловленном требованиями к точности решения исходной задачи.

Существующие беспереборные процедуры анализа упорядоченных данных на основе принципа динамического программирования позволяют учитывать лишь весьма узкий класс априорных предположений о решаемой прикладной задаче, оставаясь в рамках линейной нормальной модели скрытого случайного поля. В то же время, специальная версия процедуры динамического программирования позволяет практически без дополнительных затрат вычислить значение критерия, максимизирующего апостериорную плотность распределения, при заданном значении степени связи для двух смежных целевых переменных.

Основная идея работы заключается в использовании для оценки степени связи каждой пары смежных переменных отношения значения целевой функции (при запрете данной паре переменных принимать различные значения) к значению данной функции при полном отсутствии такой связи. Вычисление данного отношения для каждой пары смежных переменных позволяет адаптивно задавать нестационарную априорную модель скрытого поля, подлежащего восстановлению, и позволяет учесть возможное наличие локальных неоднородностей и скачков.

Работа поддержана грантом РФФИ № 13-07-00529.

- [1] *Грачева И. А., Копылов А. В.* Адаптивный параметрический алгоритм обработки изображений // Машинное обучение и анализ данных, 2013, № 6. <http://jmla.org/papers/doc/2013/no6/>.

## Ансамблевые алгоритмы сегментации спутниковых изображений по спектральным и пространственным признакам

*Пестунов Игорь Алексеевич*<sup>1</sup>

*Мельников Павел Владимирович*<sup>2</sup>

*Синявский Юрий Николаевич*<sup>3</sup>

<sup>1</sup>pestunov@ict.nsc.ru, <sup>2</sup>pvm96@yandex.ru, <sup>3</sup>yorikmail@gmail.com

<sup>1,2,3</sup>Новосибирск, Институт вычислительных технологий СО РАН

Сегментация является одним из важнейших этапов анализа цифровых изображений. Она заключается в разбиении изображения на непересекающиеся области на основе однородности (похожести) их спектральных и пространственных (текстура, форма, размер и др.) характеристик. В работе [1] предложен вычислительно эффективный алгоритм сегментации изображений EMeanSC по спектральным признакам, основанный на ансамбле непараметрических алгоритмов кластеризации. Как показали исследования, он обеспечивает высокое качество сегментации и устойчивость получаемых результатов по отношению к вариациям значений настраиваемых параметров. Алгоритм EMeanSC учитывает лишь спектральные признаки, поэтому картосхемы, получаемые в ходе сегментации спутниковых изображений, являются чрезмерно раздробленными.

Для снижения раздробленности и повышения качества картосхем в докладе предлагается применить алгоритм морфологической сегментации с использованием пространственной информации, содержащейся в изображении. Алгоритм основан на построении минимального остовного леса графа изображения. Вершинами графа являются пиксели изображения, а ребра строятся с учетом пространственного расположения пикселей (веса ребер равны значению выбранной меры схожести пикселей). Минимальный лес строится в соответствии с набором маркеров, выбранных на картосхеме, которая получена алгоритмом EMeanSC. Для получения устойчивого результата, не зависящего от выбора маркеров, при морфологической сегментации строится ансамбль алгоритмов, в каждом из которых используется свой случайный набор маркеров. Итоговая картосхема получается путем отнесения каждого пикселя к кластеру, получившему в процессе реализации ансамбля морфологических алгоритмов наибольшее количество голосов. В докладе приводятся результаты обработки изображений, полученных со спутников Alos и Ikonos, подтверждающие высокое качество получаемых картосхем.

Работа поддержана грантами РФФИ № 11-07-00202 и 11-07-00346.

- [1] *Пестунов И.А., Бериков В.Б., Синявский Ю.Н.* Сегментация многоспектральных изображений на основе ансамбля непараметрических алгоритмов кластеризации // Вестник СибГАУ. 2010. Т. 31, № 5. С. 45–56.

## Анализ видеопоследовательностей, формируемых капилляроскопом

*Чочиа Павел Антонович*

chochia@iitp.ru

Москва, Институт проблем передачи информации РАН

Рассматриваются методы автоматической обработки и анализа видеопоследовательностей, регистрируемых компьютерным капилляроскопом — устройством для неинвазивного исследования и диагностики микроциркуляторного кровяного русла. Излагаются алгоритмы преобразования видеоданных, служащие для извлечения необходимой информации о параметрах капилляров — статических (размеры, форма) и динамических (скорость кровотока). Такие параметры позволяют врачу делать выводы о состоянии здоровья пациента и его предрасположенности к различным заболеваниям.

На этапе обработки осуществляется фильтрация, пространственная синхронизация и предварительное преобразование кадров видеопоследовательности. На основе корреляционного совмещения кадров компенсируются колебания и устраняется дрейф наблюдаемых объектов. Суммированием множества кадров формируется обобщенное интегральное изображение [1], являющееся основой для последующего анализа.

Анализ интегрального изображения позволяет обнаружить сосуды и выбрать среди них мажоритарный капилляр. Для этого формируется карта контурных линий капилляров, осуществляется синтаксический анализ, фильтрация и упрощение полученной карты. Линии переводятся в векторное представление и составляются морфологические характеристики, на основе которых можно определять форму контурных линий и размеры капилляра [2].

Найденное векторное представление контурных линий дает возможность получить гладкое отображение области капилляра, форма которого не описывается аналитически, в область простой формы. Строится соответствующее геометрическое преобразование, переводящее обнаруженную область в прямоугольную, которое выполняется для каждого кадра видеопоследовательности. Это позволяет привести исходные данные к виду, удобному для анализа динамики кровотока. Далее уже могут определяться скорость движения, объем протекающей крови и выполняться другие измерения.

- [1] Чочиа П. А. Предварительная обработка видеопоследовательностей, формируемых капилляроскопом // Информационные процессы, 2011, Т. 11, № 1. — С. 76–85. <http://www.iip.ru/2011/76-85-2011.pdf>.
- [2] Чочиа П. А. Обнаружение капилляров на изображениях, формируемых капилляроскопом // Информационные процессы, 2013, Т. 13, № 2. (в печати).

## Динамическая сегментация пары кадров

*Хашин Сергей Иванович*

khash2@mail.ru

Иваново, Ивановский Государственный Университет

В объектно-ориентированных алгоритмах сжатия видеоданных важнейший этап – сегментация текущего кадра. От его качества зависит и общая эффективность сжатия. Возможности обычных методов сегментации в этом направлении практически исчерпаны. Но при кодировании видео мы имеем на входе не только один предыдущий кадр, а цепочку предыдущих кадров. Это позволяет радикально улучшить качество сегментации.

Рассмотрим пару искусственных кадров: фон и четыре спрайта. И для фона, и для каждого спрайта задано свое аффинное преобразование. В результате динамической сегментации получается практически точно исходная сегментация, ее погрешность оказывается равной 8.6.



Рис. 1. Кадр со спрайтами и его сегментация

В реальных видеопоследовательностях различного размера, от CIF до HD, получается от 2 до 15 сегментов, погрешность сегментации — от 2.6 до 11.2.

Таким образом, типичное количество сегментов, требуемых для приемлемой погрешности уменьшается с 200–500 до 5–15. Основной составляющей погрешности является не ошибки сегментации и нахождения движения, а «uncovered background» и цветовое различие между кадрами, то есть причины, которые невозможно устранить выбором сегментации и аффинных преобразований.

Работа поддержана грантом РФФИ № 11-07-00653.

- [1] *Хашин С.И.* Динамическая сегментация последовательности кадров // Машинное обучение и анализ данных, <http://jmla.org/papers/index.php/JMLDA/author/submission/40>.

## Метод автоматической сегментации человеческих фигур на изображениях

*Малин Иван Константинович*<sup>1</sup>

*Вишняков Борис Ваисович*<sup>2</sup>

<sup>1</sup>imalin@gosniias.ru, <sup>2</sup>vishnyakov@gosniias.ru

Москва, <sup>1,2</sup>ФГУП ГосНИИАС

Автоматическая сегментация человеческих фигур на изображениях является одной из наиболее важных задач машинного зрения. Можно выделить два основных подхода к данной проблеме: методы, основанные на процедуре, аналогичной голосованию Хафа, и методы, основанные на поиске с использованием скользящего окна. Методы первой группы (Leibe et al., 2005; Gall et al., 2009) опираются на построение модели пространственного распределения визуальных особенностей объекта. При поиске объекта визуальные особенности изображения голосуют с некоторым весом за возможное расположение центра искомого объекта. Методы с использованием скользящего окна (Dollar et al., 2008; Dalal et al., 2005; Jones et al., 2008) позволяют с некоторой вероятностью определить, является ли данный фрагмент кадра изображением человека. Перебор всевозможных фрагментов позволяет локализовать людей на изображении.

В работе (Dollar, 2008) был предложен метод интегральных каналов признаков, использующий в качестве входных данных разнородные каналы изображения: непосредственно яркость, направление и абсолютное значение градиента изображения и пр. При этом используется только информация со статичного изображения. В то же время движение объекта несет в себе дополнительную информацию, потенциально позволяющую улучшить точность и надежность сегментации.

Таким образом, целью данной работы является исследование возможности использования дополнительных каналов, несущих информацию о движении объектов. В качестве таких каналов были выбраны составляющие кратнорегрессионного псевдоспектра. Также рассматривается характер используемых признаков. В оригинальном методе интегральных каналов признаков каждый признак представляет собой сумму значений одного канала в некоторой прямоугольной области. В данной работе исследована применимость классических хаароподобных признаков в методе интегральных каналов.

Работа поддержана грантом РФФИ № 12-07-00798-а.

- [1] *Малин И. К., Вишняков Б. В.* Метод автоматического выделения человеческих фигур на изображениях с использованием слоев кратнорегрессионного спектра в качестве интегральных каналов признаков // ВКИТ, М.: Спектр, 2013. Принята к публикации.

## Модель статического фона с использованием дескрипторов LBP и LTP

*Вишняков Борис Ваисович*<sup>1</sup>

*Малин Иван Константинович*<sup>2</sup>

*Егоров Антон Александрович*<sup>3</sup>

<sup>1</sup>vishnyakov@gosniias.ru, <sup>2</sup>imalin@gosniias.ru,

<sup>3</sup>aegorov@gosniias.ru

Москва, <sup>1,2,3</sup>ФГУП ГосНИИАС

В задачах машинного зрения, связанных с выделением движущихся объектов или нахождением изменившихся областей сцены, основной проблемой является построение модели статического фона.

Основными современными моделями фона являются Mixture of Gaussians (MoG) [Stauffer, Grimson, 1998] и Kernel Density Estimator (KDE) [Mittal, Paragios, 2004]. Также в статье [Вишняков и др., 2007] была предложена быстрая модель фона на основе вычисления кратнорегрессионных псевдоспектров. Однако все вышеперечисленные методы, обладая своими плюсами, имеют один общий недостаток: они используют накопленную информацию лишь в отдельном пикселе за конечный промежуток времени. При этом взаимовлияние соседних пикселей никак не учитывается. В последнее время стали появляться работы, в которых авторы пытаются подменить отдельно взятый пиксель в моделях MoG или KDE дескриптором его окрестности [Liao et al, 2010]. Такой подход значительно улучшает качественные характеристики как используемой модели фона, так и алгоритмов видеоаналитики. Однако в силу своей вычислительной сложности он оказывается практически неприменимым для обработки кадров с нескольких камер в режиме реального времени.

В данной работе строится модель фона, основанная на вычислении кратнорегрессионных спектров LBP и LTP-гистограмм. Вместо оценки значения отдельного пикселя изображения с помощью процесса авторегрессии предлагается вычислять оценку отдельных компонент дескрипторов LBP или LTP. Данный подход позволяет стабилизировать значение дескриптора, а также получить модель фона, которая будет практически нечувствительна к изменениям яркости в сцене.

Основа в виде адаптивной модели фона позволяет создать робастный алгоритм видеонаблюдения, способный работать в режиме реального времени, который справляется с изменениями яркости, шумами матрицы и изменениями в самой сцене.

Работа поддержана грантом РФФИ № 12-07-00798-а.

- [1] Вишняков Б. В., Малин И. К., Егоров А. И. Построение модели статического фона с помощью авторегрессионной фильтрации дескрипторов LBP и LTP // ВКИТ, М.: Спектр, 2013. Принята к публикации.

## Конкурс Фонда перспективных исследований на лучший прототип технологии распознавания образов

*Гарбук Сергей Владимирович*

garbuk@list.ru

Москва, Фонд перспективных исследований

Фонд перспективных исследований создан в 2012 году на основании Федерального закона ФЗ-174 для поддержки научных исследований и разработок, связанных с высокой степенью риска достижения качественно новых результатов в военно-технической, технологической и социально-экономических сферах.

Одним из приоритетных направлений Фонда является создание прорывных решений в области искусственного интеллекта, в частности, в задачах распознавания. С этой целью Фонд организует конкурс на лучший прототип технологии распознавания образов.

В рамках конкурса участникам предлагается разработать алгоритмы распознавания объектов на изображениях, моделирующие интеллектуальную деятельность мозга человека. Под этим понимается создание когнитивных систем, способных к продолжительному автономному обучению, обобщению приобретённых знаний и пониманию контекста многозначной информации для решения комплексных проблем предметной области. Для рассмотрения предлагаются следующие задачи:

- распознавание людей, автомобилей, зданий и сооружений, дорог по фотоснимкам земной поверхности с БПЛА и космических систем ДЗЗ;
- распознавание лиц людей в различном возрасте, в различных эмоциональных состояниях, среди лиц похожих людей, в т. ч. близнецов.

Фондом планируется проведение конкурса в несколько этапов, цель которых — стимулировать фундаментальные сдвиги в данном направлении, а также инициировать создание научной базы исследований и разработок, необходимых для трансформационного прорыва технологий распознавания образов.

Каждый из этапов конкурса завершается показом достижений на площадке Фонда, по итогам которого экспертным советом Фонда решается вопрос о создании научно-исследовательских лаблетов по разработке и синтезу нового знания в области распознавания. Функционирование данных лаблетов будет осуществляться под руководством Фонда в рамках многолетней программы по развитию и внедрению научных достижений и технических разработок в производство. Помимо прочего, коллективы, продемонстрировавшие лучшие результаты, будут награждены денежными призами.

Информация о конкурсе доступна по адресу <http://fpi-contest.ru>.

## Отбор и обучение бинарных классификаторов в задаче распознавания лиц

*Горбацевич Владимир Сергеевич*

gvs@gosniias.ru

Москва, ФГУП ГосНИИАС

В работе рассматривается задача построения, отбора и обучения бинарных классификаторов (БК) для биометрического распознавания лиц. Основной идеей работы является использования в качестве элементов бинарного вектора признаков пороговых локальных бинарных паттернов (ЛБП), вычисляемых для вейвлет-разложения изображения лица. Стандартный ЛБП имеет вид

$$C_{LBP}(x, y) = [I(x, y) - I(x + dx, y + dy) \geq 0],$$

где  $I(x, y)$  — функция яркости,  $dx, dy$  — смещение точки окрестности. Пороговый ЛБП (ПЛБП) имеет дополнительный параметр  $\theta$ :

$$C_{LBP}(x, y, \theta) = [I(x, y) - I(x + dx, y + dy) \geq \theta].$$

Значения порогов каждого БК определяются в результате обучения, призванного обеспечить одинаковую реакцию БК на «свои» изображения (одного лица) и различную на «чужие» (разных лица). При этом важны не выходные значения БК (0 или 1), а их совпадение или несовпадение. Такая нестандартная постановка задачи обучения с учителем дополнительно осложняется большим количеством обучаемых БК, большими размерами обучающей выборки и наличием всего двух экземпляров изображений каждого класса.

В работе предлагается метод обучения таких классификаторов, а также метод отбора наиболее информативных БК. Значения порога каждого БК определяется соотношением весов ошибок 1-го и 2-го рода. В ходе обучения эти соотношения итеративно изменяются из-за изменения весов «своих» и «чужих» сравнений в обучающей выборке в зависимости от результатов их текущего распознавания (подобно взвешиванию в алгоритме Adaboost). Затем отбор признаков осуществляется путем выбора заданного числа «лучших» БК на каждой итерации.

Проведено первичное сравнительное тестирование метода на базе изображений colorferet (fa-fb). Получены сравнимые с конкурентными методами вероятности распознавания при коротком шаблоне поиска (до 600 байт) и высокой скорости сравнения (до 10 млн. шаблонов в секунду).

Работа поддержана грантом РФФИ № 11-08-01039.

- [1] *Горбацевич В. С.* Отбор и обучение бинарных классификаторов в задаче распознавания лиц // Вестник компьютерных и информационных технологий. <http://www.vkit.ru/>. (принята к рассмотрению)

## Определение точной границы зрачка

*Чинаев Николай Николаевич*<sup>1</sup>

*Матвеев Иван Алексеевич*<sup>2</sup>

<sup>1</sup>chinaev.n@gmail.com, <sup>2</sup>matveev@ccas.ru

Москва, <sup>1</sup>МФТИ, <sup>2</sup>ВЦ РАН

Предложен метод выделения контура зрачка на изображении глаза. Контур строится из точек, лежащих на границе области зрачка, полученной при бинаризации изображения как компонента связности множества тёмных точек. Для выбора правильной (наиболее близкой к окружности) компоненты (или её части) используется преобразование Хафа для граничных точек, приводящее к выделению центра её округлой части. Радиус окружности определяется как максимум гистограммы расстояний граничных точек до выделенного центра. Сами точки контура зрачка затем выбираются из всего множества точек границы как имеющие близкое к этому радиусу расстояние до центра.

Проведена апробация работы алгоритма на наборе изображений из открытой базы данных ВАН. Более 95% изображений имеют отклонения центра зрачка от истинного положения меньше 10 пикселей и отклонения радиуса меньше 5 пикселей.

Использование преобразования Хафа позволяет корректно выделять истинные центр, радиус и контур зрачка даже при наличии больших помех, при условии лишь частичной видимости контура зрачка на изображении. Недостатком метода является потенциально большое время работы, возникающее при переборе многих компонент связности и нескольких порогов бинаризации. Этот недостаток частично скомпенсирован введением признака качества компоненты связности.

Работа поддержана грантом РФФИ № 12-07-00778.

- [1] *Чинаев Н. Н., Матвеев И. А.* Определение точной границы зрачка // Машинное обучение и анализ данных, <http://jmla.org/papers/doc/2013>.

## Распознавание жестов ручных азбук глухонемых

*Нагапетян Ваагн Эдвардович*

vahagnahapetyan@gmail.com

Москва, Российский университет дружбы народов

В настоящей работе предлагается метод и разработанная программная система автоматического распознавания жестов ручных азбук ASL (American Sign Language) и русского жестового языка.

В качестве устройства ввода информации о жестах, отображающих цифры и буквы, выступает трёхмерный сенсор нового поколения Asus Xtion Pro Live. Выбранный сенсор позволяет со скоростью 30 кадров в секунду получить дальностное изображение человека, которая представляет собой облако точек. Посредством применения программных библиотек OpenNI и NITE в облаке точек ищется точка ладони человека. Вокруг найденной точки строится сфера и все точки, которые не входят в эту сферу удаляются. Оставшейся точки разделяются на связанные компоненты, из которых удаляются те, которые не содержат точку ладони. В итоге остаются точки ладони, координаты абсцисс и ординат которой показывают соответственно столбец и строку в дальностном растровом изображении, а третья координата – глубину, т.е. расстояние до сенсора. Полученные точки проецируют на плоскость  $XU$ , получая тем самым двумерную фигуру ладони и пальцев человека. Строится геометрический скелет фигуры ладони, который в дальнейшем выступает как дескриптор статического жеста руки. Идентификация статического жеста осуществляется посредством сравнения скелета показанного жеста с эталонными скелетами. Расстояние между двумя скелетами оценивается посредством алгоритма динамической трансформации шкалы времени (Dynamic Time Warping).

Из проведённых экспериментов видно, что средняя точность распознавания для статических жестов латинских букв составила 84.08%, полнота — 81.2%. Те же показатели для русских букв составили соответственно 83.4% и 76.7%.

Распознавание динамических жестов ручных азбук осуществляется посредством отслеживания позиции и конфигурации ладони в кадрах видеоряда. На основе этих данных видеоряд разделяется на отдельные сегменты, где каждый сегмент соответствует показу отдельного жеста. Жест идентифицируется посредством сравнения траектории движения и конфигурации ладони с соответствующими ключевыми характеристиками эталонных жестов.

- [1] *Нагапетян В. Э.* Распознавание жестов ручной азбуки ASL // Вестник Российского университета дружбы народов. Серия: математика, информатика, физика, М.: Российский университет дружбы народов, 2013. — С. 105–113.

## Метод целевого нахождения согласованных описаний при распознавании объектов изображений

*Лебедев Леонид Иванович*

lebedev@pmk.unn.ru

Нижегород, НИИ прикладной математики и кибернетики

Нижегородского государственного университета им. Н. И. Лобачевского

Распознавание объектов изображений  $\mathbf{O}$  с использованием КЭКМ осуществляется на основе нахождения оценок сходства их с эталонами  $\mathbf{E}_i$ , которые вычисляются на базе согласованных описаний. Ранее было показано, что задача получения согласованных описаний может быть сведена к задаче нахождения глобального минимума функции оценки близости  $\varepsilon_m(\mathbf{E}_i, \mathbf{O}, s)$ , где  $s$  — параметр описания контура объекта. В отличие от ранее разработанных итеративных алгоритмов предлагаемый метод дает точное решение задачи нахождения местоположения глобального минимума. В основе этого метода лежит следующая теорема. Пусть контур  $\mathbf{C}_{\mathbf{E}_i}$  задается описанием эталона  $\mathbf{E}_i$ . Обозначим  $\mathbf{K}(\mathbf{C}_{\mathbf{E}_i}, \mathbf{G})$  класс эквивалентности, порожденный эталоном  $\mathbf{E}_i$  и преобразованием  $\mathbf{G}$ , представляющим аффинную группу Ли на плоскости.

**Теорема 1.** *Если эталон  $\mathbf{E}_i$  и объект  $\mathbf{O}$  принадлежат одному и тому же классу эквивалентности  $\mathbf{K}(\mathbf{C}_{\mathbf{E}_i}, \mathbf{G})$ , то местоположение глобального минимума функции  $\varepsilon_m(\mathbf{E}_i, \mathbf{O}, s)$  можно определить аналитически на основе вычислений ее значений в заданном (определенным образом) конечном числе точек (метод целевого нахождения).*

Из доказательства теоремы следует, что если описания являются согласованными, то при наложении контуров хотя бы одна из точек эталона совпадет с какой-либо точкой объекта. Отсюда следует, что если исходные описания контуров эталона и объекта содержат  $m$  и  $k$  точек соответственно, то из бесконечного числа циклических описаний объекта для решения задачи согласования необходимо проверить только  $mk$  из них, то есть проверить на согласованность все описания объекта, когда каждая точка контура эталона совпадает с каждой точкой объекта.

Нетрудно видеть, что вычислительная сложность анонсируемого метода даже выше, чем у итеративных алгоритмов определения местоположения глобального минимума. Понизить сложность метода можно, если найти способ отбора незначительного количества описаний из  $mk$ , для которых необходимо вычислять оценки близости  $\varepsilon_m(\mathbf{E}_i, \mathbf{O}, s)$ .

Работа поддержана грантом РФФИ № 13-07-00521.

- [1] *Лебедев Л. И.* Корреляционно-экстремальные контурные методы распознавания. Теоретические основы // Учебное пособие, Нижний Новгород: Издво Нижегородского университета, 2013, — 114 с.

## Восстановление симметричности точек на изображениях объектов с отражательной симметрией

*Каркищенко Александр Николаевич*<sup>1</sup>

*Мнухин Валерий Борисович*<sup>2</sup>

<sup>1</sup>karkishalex@gmail.com, <sup>2</sup>mnukhin.valeriy@mail.ru

<sup>1</sup>Ставрополь, Северо-Кавказский Федеральный Университет,

<sup>2</sup>Таганрог, Южный Федеральный Университет

В последнее время многочисленные усилия при анализе формы объектов, заданных в оцифрованном виде, были сосредоточены на выявлении симметрии. Как правило, на «атомарном» уровне применяемые при этом методы сводятся к исследованию симметричности так называемых характерных точек, смысл которых зависит от решаемой задачи.

В данной работе предлагается ряд методов, которые позволяют уточнить положение характерных точек, описывающих объекты на изображении, на основе априорной информации об их симметричном расположении. Суть проблемы состоит в том, что локализация точек на изображении всегда осуществляется с некоторой погрешностью, величина которой зависит от многих факторов — используемого алгоритма локализации, качества изображения в целом, зашумления в области, содержащей локализуемую точку и др. Как следствие, координаты обнаруженных точек, про которые априори известно, что они симметричны, в реальности этому условию не удовлетворяют. Поэтому целесообразно воспользоваться известной информацией о симметричности точек для уточнения их положения. При этом само уточнение необходимо осуществлять в некотором смысле наилучшим образом, скажем, симметричность должна достигаться минимальным изменением их положения.

Одной из наиболее типичных задач, где могут применяться данные методы уточнения «по симметрии», являются задачи биометрического распознавания, в которых правильное определение характерных точек является критическим фактором успешного решения задачи. Здесь достаточно указать на то, что точность детекции и распознавания лиц существенно зависит от точности определения центров зрачков на лице. При этом в задачах идентификации личности, основанных на определении положения характерных точек, заведомо присутствует фактор симметричности, обусловленный зеркальной симметричностью формы лица анфас.

Работа поддержана грантами РФФИ № 11-07-00591 и № 13-07-00327.

- [1] Каркищенко А. Н., Мнухин В. Б. Восстановление симметричности точек на изображениях объектов с отражательной симметрией // Машинное обучение и анализ данных, 2013. — С. 1–9. <http://jmla.org/papers/doc/2013/no6>.

## Обнаружение трехмерных групп точечных объектов на основе анализа графа иерархической группировки объемной сцены

*Кревецкий Александр Владимирович*

krevetskyav@volgatech.net

Йошкар-Ола, Поволжский государственный технологический университет

В современных системах технического зрения появилась возможность формирования трехмерных изображений пространственных объектов. Особый тип образов из несвязных точечных отметок получил название пространственных групповых точечных объектов (ГТО). ГТО может быть образован отметками малоразмерных объектов или характерными точками сплошных (связных) объектов. Поскольку под воздействием мешающих факторов возможны ошибки первого и второго родов при формировании ГТО, то перед их распознаванием необходимо вначале обнаружить, разрешить и пространственно ограничить ГТО (далее для краткости просто «обнаружить»).

Самые устойчивые к помехам в виде ложных отметок и пропуску части полезных известные методы обнаружения ГТО строятся на анализе плотности точек как некоторого их числа в области заданной формы. Такой подход оказывается недостаточно эффективным, когда форма данной области и ее ориентация заранее неизвестны.

Предлагаемый в данной работе метод обнаружения ГТО базируется на байесовском подходе анализа графа иерархической группировки объемной сцены с точечными отметками [1].

Результатами исследования стали теоретические законы распределения длин ребер минимального дерева и числа связываемых ребрами не более пороговой длины точечных отметок трехмерного пуассоновского потока. В работе получены формулы для определения порогов по длине ребра и числу связанных короткими ребрами точечных отметок, оптимальные по критерию Неймана-Пирсона. Приводятся теоретические и эмпирические характеристики обнаружения пространственно компактных ГТО (в том числе с цепочечной и облачной структурой) полученным методом.

Предлагаемый метод обнаружения пространственно компактных ГТО инвариантен к их форме и позволяет разрешать и пространственно ограничивать ГТО при априорной неопределенности относительно ракурса наблюдения.

Работа выполнена в рамках проекта РФФИ № 13-01-00427.

- [1] *Кревецкий А. В.* Инвариантные к форме обнаружение и пространственная локализация групп точечных объектов в трехмерном пространстве // Вестник МарГТУ, Йошкар-Ола: МарГТУ, 2011. №1. — С. 47–53.

## Комбинирование одноклассовых классификаторов в задачах сегментации изображений, содержащих объекты топологии печатной платы

*Ларин Александр Олегович*<sup>1</sup>

*Середин Олег Сергеевич*<sup>2</sup>

*Кондрашов Владимир Владимирович*<sup>3</sup>

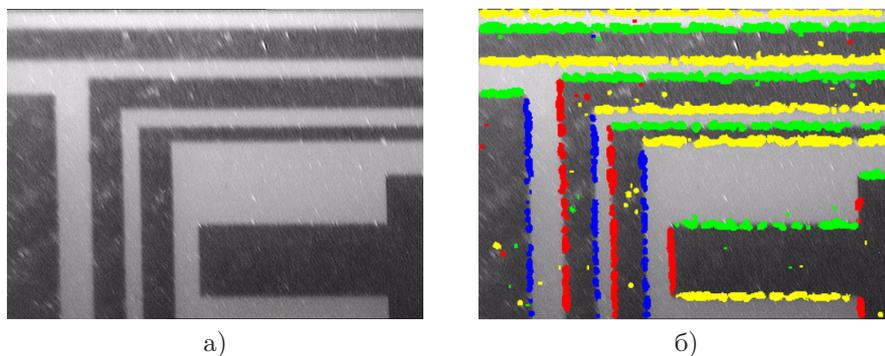
<sup>1</sup>ekzebox@gmail.com, <sup>2</sup>oseredin@yandex.ru

Москва, <sup>1</sup>Московский физико-технический институт,

Тула, <sup>2,3</sup>Тульский государственный университет

Решение задачи автоматизации установок лазерной подгонки резистивных элементов требует предварительной сегментации элементов топологической структуры печатной платы на основе изображения, полученного с видеокамеры.

Задача заключается в том, что на изображениях, полученных с видеокамеры, необходимо выделить границы основных элементов печатной платы: дорожек и резисторов. Применение стандартных алгоритмов машинного зрения может быть затруднено спецификой изображений платы. В работе мы предлагаем подход сегментации изображений, основанный на комбинировании одноклассовых классификаторов Такса в пространстве признаков LBP (Local Binary Patterns).



**Рис. 1.** Пример исходного изображения печатной платы (а) и результат применения к нему алгоритма сегментации, основанного на комбинировании одноклассовых классификаторов (б).

Работа поддержана грантом РФФИ № 12-07-92000-ННС\_а.

- [1] *Ларин А. О.* Комбинирование одноклассовых классификаторов в задачах сегментации изображений, содержащих объекты топологии печатной платы // *Машинное обучение и анализ данных*, 2013. — принято в печать. <http://jmla.org>.

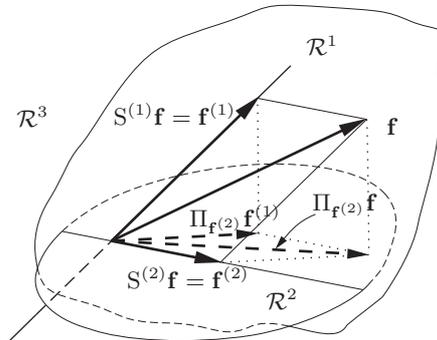
## Косые проекторы и относительные формы в морфологии изображений

*Пытьев Юрий Петрович*

yuri.pytyev@gmail.com

Москва, МГУ им. М. В. Ломоносова

В докладе рассматриваются [1] новые математические методы морфологического анализа классов изображений и сравнительного анализа их форм как инвариантных относительно условий регистрации изображений носителей информации об изображенных сценах, объектах и т. п. Рассматриваются понятия абсолютных и относительных форм классов изображений и представляющих их ортогональных и косых проекторов, позволяющие охарактеризовать морфологические зависимости относительных форм значениями индекса морфологической независимости и абсолютных форм значениями индекса морфологической связности. Рассматриваются основанные на технике косого проецирования методы решения задач сравнительного анализа абсолютных и относительных форм классов изображений, морфологической фильтрации и идентификации изображений, выделения неизвестных объектов на изображении и др. Суть техники косого проецирования иллюстрирует рис. 1.



**Рис. 1.** Косые проекторы  $S^{(i)}: \mathcal{R}^3 \rightarrow \mathcal{R}^3$ ,  $i = 1, 2$ ,  $S^{(1)}$  проецирует на  $\mathcal{R}^1$  вдоль  $\mathcal{R}^2$ ,  $S^{(2)} = I - S^{(1)}$  проецирует на  $\mathcal{R}^2$  вдоль  $\mathcal{R}^1$ ,  $\mathcal{R}^3 = \mathcal{R}^1 \oplus \mathcal{R}^2$ . Если  $\mathcal{R}^1$  — форма сигнала  $\mathbf{f}^{(1)}$ ,  $\mathcal{R}^2$  — форма помехи  $\mathbf{f}^{(2)}$ ,  $\Pi_{\mathbf{f}^{(i)}}$  ортогонально проецирует на  $\mathcal{R}^i$ ,  $i = 1, 2$ , и  $\mathbf{f} = \mathbf{f}^{(1)} + \mathbf{f}^{(2)}$  — наблюдаемый сигнал, то  $S^{(1)}\mathbf{f} = \mathbf{f}^{(1)}$ ,  $S^{(2)}\mathbf{f} = \mathbf{f}^{(2)}$ , в то время как  $\Pi_{\mathbf{f}^{(1)}}\mathbf{f} = \Pi_{\mathbf{f}^{(1)}}\mathbf{f}^{(1)} + \Pi_{\mathbf{f}^{(1)}}\mathbf{f}^{(2)} = \mathbf{f}^{(1)} + \Pi_{\mathbf{f}^{(1)}}\mathbf{f}^{(2)}$ ,  $\Pi_{\mathbf{f}^{(2)}}\mathbf{f} = \Pi_{\mathbf{f}^{(2)}}\mathbf{f}^{(1)} + \Pi_{\mathbf{f}^{(2)}}\mathbf{f}^{(2)} = \Pi_{\mathbf{f}^{(2)}}\mathbf{f}^{(1)} + \mathbf{f}^{(2)}$  суть наиболее точные приближения  $\mathbf{f}$  сигналами  $\mathbf{f}^{(1)} \in \mathcal{R}^1$  и помехами  $\mathbf{f}^{(2)} \in \mathcal{R}^2$ , используемые в стандартном морфологическом анализе сигналов.

Работа поддержана грантом РФФИ № 11-07-00722.

- [1] *Пытьев Ю. П.* Косые проекторы и относительные формы в морфологии изображений // ЖВМ и МФ, М.: Pleiades Publishing, 2013. — № 12.

## Морфологический метод выделения неизвестного объекта на изображении, полученном при неизвестном освещении

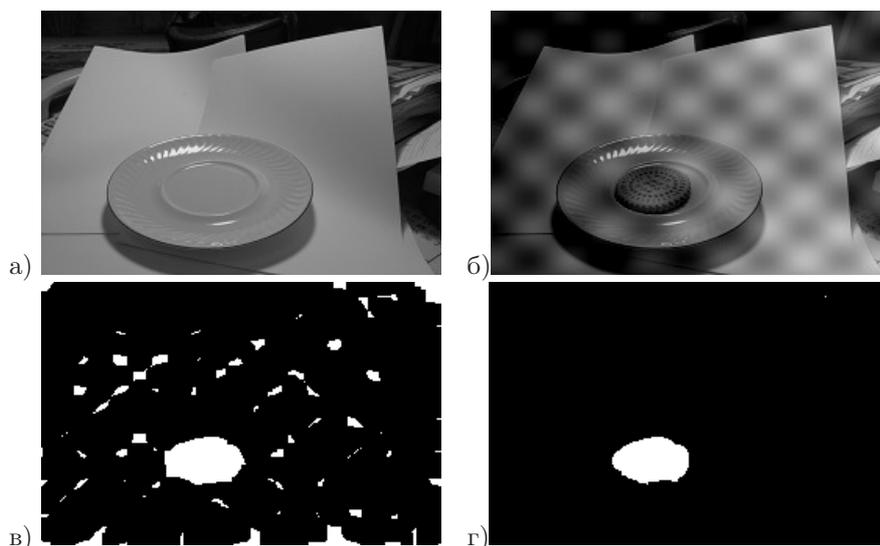
*Пытьев Юрий Петрович<sup>1</sup>*

*Нагорный Юрий Михайлович<sup>2</sup>*

<sup>1</sup>yuri.pytyev@gmail.com, <sup>2</sup>imtest@yandex.ru

Москва, МГУ им. М. В. Ломоносова

В докладе рассматривается вариант морфологического метода выделения появившегося неизвестного объекта на известном фоне при неконтролируемом изменении освещения (при появлении «градиентов», «волн», «пятен» и т. п. в освещении), предложенный метод сравнивается со стандартным морфологическим методом решения этой же задачи.



**Рис. 1.** а) изображение «известного фона»; б) изображение, предъявленное для анализа (появился объект, изменились условия освещения); в) результат выделения появившегося объекта стандартным морфологическим методом; г) результат решения этой же задачи предложенным морфологическим методом.

[1] *Пытьев Ю. П., Нагорный Ю. М.* Новые морфологические методы выделения объектов и фильтрации // (подг. к печати в ЖВМ и МФ).

## Морфологическое сравнение образов по сложности

*Визильтер Юрий Валентинович*<sup>1</sup>

*Рубис Алексей Юрьевич*<sup>2</sup>

<sup>1</sup>viz@gosniias.ru, <sup>2</sup>arcelt@mail.ru

Москва, ФГУП «Государственный научно-исследовательский институт авиационных систем» (ГосНИИАС)

В математической морфологии Пытьева используется понятие «морфологической сложности» формы изображения, описываемое отношением частичного порядка «более простой/сложный по форме». В данной работе предлагается способ корректного распространения данного отношения на множество всех мозаичных форм по аналогии с отношением «более прогрессивный вид» в эволюционной биологии. Поскольку у любых двух видов на эволюционном дереве имеется ближайший общий предок, тот вид, который относительно ближе к этому общему предку, считается относительно более «примитивным», а тот, который дальше от него — относительно более «прогрессивным». Подобным образом, на основе дерева относительного «усложнения/упрощения», можно сравнивать по сложности формы в морфологии Пытьева. Для оценки удаленности форм от общего «предка» предложено использовать метрику оценки геометрических отличий (ОГО-метрику). При этом использование дерева сложности форм и соответствующей абсолютной меры сложности позволяет переопределить ОГО-метрику как расстояние на дереве сложности.

Показано, что описанная схема построения набора характеристик, связанных с морфологической сложностью, может быть распространена на случай любых проективных морфологий. Эта обобщенная схема включает: абсолютную меру сложности образов; элементарные преобразования образов, монотонно изменяющие эту меру; решетку образов с соответствующими отношениями частичного порядка и операциями взятия супремума и инфимума образов по сложности; метрику сравнения образов, основанную на дереве и мере сложности; меру относительной сложности образов, согласованную с абсолютной мерой сложности.

Такую схему предложено называть аппаратом морфологического сравнения образов по сложности. Приведен пример ее реализации для морфологии Серра.

Работа поддержана грантами РФФИ 12-07-31186-мол-а, 13-08-01071-а.

- [1] *Визильтер Ю. В., Рубис А. Ю.* Морфологическое сравнение образов по сложности // Вестник компьютерных и информационных технологий, Москва: Издательский дом «Спектр», 2013 (принято в печать)

## Параметрические и морфологические спектры

*Визильтер Юрий Валентинович*<sup>1</sup>

*Сидякин Сергей Владимирович*<sup>2</sup>

<sup>1</sup>viz@gosniias.ru, <sup>2</sup>sersid@bk.ru

Москва, <sup>1,2</sup> ФГУП «Государственный научно-исследовательский институт авиационных систем» (ГосНИИАС)

В работе П. Марагоса (Maragos, 1989) по аналогии со спектрами Фурье был предложен способ описания фигур и изображений при помощи форморазмерных спектров, вычисляемых средствами математической морфологии Серра. В последующие годы многими исследователями, включая авторов данной статьи, был предложен ряд дескрипторов изображений, построенных в той или иной степени по аналогии со спектрами Марагоса и также названных спектрами.

В данной работе предлагается обобщающий формализм «параметрических спектров» для описания всех спектров такого типа с единых позиций. Параметрический спектр представляет собой плотность распределения некоторой меры (численной характеристики) образа (изображения) по анализируемому параметру, от которого эта мера монотонно зависит.

Предложенный формализм позволяет ввести классификацию возможных параметрических спектров. Наиболее общим случаем параметрических спектров являются спектры точности измерения от параметра разрешения. В качестве частных случаев они включают: спектры на основе операторов описания, спектры на основе операторов фильтрации, спектры на основе частичного порядка и монотонных операторов фильтрации, спектры на основе вложенных проекторов и расстояния, спектры на основе нормированных линейных пространств.

Отдельно рассмотрен подкласс морфологических спектров, представляющих собой спектры точности реконструкции по параметру сложности описания. Этот подкласс включает, в частности, спектры на основе операторов описания со структурирующим параметром и спектры на основе фильтрации, состоящей из последовательных операций описания и реконструкции.

Показана связь параметрических спектров с соответствующими параметрическими разложениями.

Обсуждается проблема устойчивости параметрических спектров при малых искажениях формы описываемых образов.

Работа поддержана грантами РФФИ 12-07-00798-а, 12-07-31218-мол-а.

- [1] *Визильтер Ю. В., Сидякин С. В.* Параметрические и морфологические спектры // Вестник компьютерных и информационных технологий, Москва: ООО Издательский дом Спектр, 2013, в печати.

## Параметрический дескриптор формы и связность базового скелета

*Жукова Ксения Владимировна*<sup>1</sup>

*Рейер Иван Александрович*<sup>2</sup>

<sup>1</sup>kz@pisem.net, <sup>2</sup>reyer@forecsys.ru

Москва, Вычислительный центр РАН

Ранее авторами была предложена концепция базового скелета — устойчивого скелетного представления формы, строящегося на основе скелета аппроксимирующей объект многоугольной фигуры. При увеличении величины точности аппроксимации базовый скелет изменяется монотонным и непрерывным образом. Ребра скелета «стираются» парами кривых — парабол и гипербол. Состав стирающей пары и положение стирающих кривых для каждого ребра определяется положением порождающих ребро элементов границы, величиной точности аппроксимации и положением определенного подмножества вершин границы. При стирании базовый скелет может разделиться на несколько связных компонент.

Монотонность и непрерывность изменения позволяют рассматривать поведение семейства базовых скелетов в целом и выбирать скелетные модели с нужной точностью аппроксимации. Кроме того, семейство базовых скелетов позволяет строить масштабируемую граничную модель формы, описывающую свойства границы, соответствующие различным степеням детализации. Был предложен параметрический дескриптор формы объекта, который можно использовать для анализа свойств формы, проявляющихся при различных значениях точности. Дескриптор представляет собой множество вершин выпуклых углов границы аппроксимирующей объект многоугольной фигуры, каждой из которых сопоставлена оценка значимости — минимальная величина точности, при которой соответствующая вершине особенность границы исключается из граничной модели. Таким образом, дескриптор дает возможность оценить количество особенностей формы, являющихся существенными при заданном значении точности, и их взаимное расположение.

В докладе рассматривается обобщение процедуры вычисления оценок значимости выпуклых особенностей для случаев нарушения связности базового скелета и исследуется возможность использования параметрического дескриптора для различных типов нарушения связности.

Работа поддержана грантами РФФИ № 11-07-00462 и № 11-01-00783.

- [1] Жукова К. В., Рейер И. А. Параметрический дескриптор формы и связность базового скелета // Машинное обучение и анализ данных, Москва: ВЦ РАН, 2013. (В печати)

## Построение функции ширины скелета бинарного изображения на основе параметрического описания многочленами Лежандра

*Кушнир Олеся Александровна*<sup>1</sup>

*Середин Олег Сергеевич*<sup>2</sup>

<sup>1</sup>kushnir-olesya@rambler.ru, <sup>2</sup>oseredin@yandex.ru

Тула, Тульский государственный университет

Доклад посвящён проблеме учёта функции ширины аппроксимированного скелетного ребра в методе сравнения скелетных графов бинарных изображений, опирающемся на процедуру парного выравнивания цепочек примитивов. Описывается прикладная задача классификации листьев лекарственных растений, для которой эта проблема актуальна.

Рассматриваются три варианта учёта функции ширины. Первый — комбинирование классификаторов, один из которых базируется на парном выравнивании цепочек скелетных примитивов, другой — на морфологических спектрах изображений. Комбинирование мер различий потребует дополнительных исследований. Второй — учет функции ширины только в вершинах аппроксимированного скелетного графа. Этот способ эффективен не для всех изображений. Третий — интерполяция функции ширины скелетного ребра полиномами Лежандра. Для этого способа были проведены экспериментальные исследования, показывающие, что коэффициенты при полиномах Лежандра каждой интерполированной функции ширины представляют собой вектор в пространстве ортогональных функций Лежандра. Несколько таких векторов, задающих функции ширины одного класса, образуют компактные сгущения.

Таким образом, коэффициенты можно использовать для сравнения форм фигур с одинаковой скелетной топологией, но различной шириной. Предлагается применить параметризованную коэффициентами Лежандра информацию о функции ширины для разработанной ранее процедуры сравнения скелетов бинарных растровых изображений, представленных цепочками примитивов.

- [1] *Кушнир О. А., Середин О. С.* Построение функции ширины скелета бинарного изображения на основе параметрического описания многочленами Лежандра // «Известия ТулГУ», серия «Естественные науки», вып. 3, Тула: Издательство ТулГУ, 2013 (принято в печать).

## Экспериментальное исследование параметров регуляризации и аппроксимации скелетных графов бинарных изображений

*Кушнир Олеся Александровна*<sup>1</sup>

*Середин Олег Сергеевич*<sup>2</sup>

*Степанов Алексей Владимирович*<sup>3</sup>

<sup>1</sup>kushnir-olesya@rambler.ru, <sup>2</sup>oseredin@yandex.ru, <sup>3</sup>tokbl@mail.ru

Тула, Тульский государственный университет

В работе систематизируются и уточняются сведения о параметрах стабилизации непрерывного скелетного графа — коэффициентах регуляризации и аппроксимации. Попутно приводится алгоритм вычисления диаметра минимальной описанной вокруг скелета окружности, являющегося масштабным множителем для применения коэффициента аппроксимации.

Проводятся экспериментальные исследования влияния коэффициентов регуляризации и аппроксимации на топологию скелета и выдвигаются предположения по определению их адекватных значений для выделения в исходном скелете его базового подграфа — устойчивого дескриптора формы фигуры при наличии шумовых изменений ее границы.

- [1] *Кушнир О.А., Середин О.С., Степанов А.В.* Экспериментальное исследование параметров регуляризации и аппроксимации скелетных графов бинарных изображений // Машинное обучение и анализ данных, 2013. [http://jmla.org/papers/index.php/JMLDA/author/submission/71/kushnir\\_2.pdf](http://jmla.org/papers/index.php/JMLDA/author/submission/71/kushnir_2.pdf).

## Иерархические структуры данных и решающие алгоритмы для классификации изображений

*Ланге Михаил Михайлович*

*Ганебных Сергей Николаевич*

lange\_mm@ccas.ru

Москва, Вычислительный центр им. А. А. Дородницына РАН

В ряде приложений, задачу распознавания образов целесообразно решать в терминах соотношения характеристик качества и вычислительной сложности, требуя минимизации вероятности ошибки при заданном ограничении на объем вычислений, либо минимизации объема вычислений при допустимой вероятности ошибки. Решение может быть получено в пространстве иерархических описаний образов с многоуровневым разрешением. Структура таких описаний удобна для сокращения объема вычислений и позволяет реализовать алгоритмы быстрого анализа геометрических форм.

В работе [1] исследуется задача распознавания образов, заданных изображениями, в пространстве древовидно структурированных представлений эллиптическими примитивами. Класс допустимых объектов, которые допускают такой способ представления, включает всевозможные объекты в виде многосвязных двумерных твердых тел с однозначно идентифицируемой системой собственных координат. Допускаются объекты в форме линейчатых (line-based) и областных (region-based) тел с полутоновой окраской. К таким объектам относятся подписи, рукописные символы, отпечатки пальцев, жесты, лица, силуэты и многие другие. Для широкого класса источников древовидные представления обладают свойством универсальности, а структура таких представлений позволяет строить сети эталонов с многоуровневым разрешением, которые допускают применение быстрых решающих алгоритмов на основе процедуры направленного поиска.

Используя многоуровневую сеть эталонов, предлагаются алгоритмы иерархического поиска решения по критерию ближайшего эталона. Получены сравнительные оценки вычислительной сложности иерархических алгоритмов относительно алгоритма полного перебора эталонов. Приведены экспериментальные зависимости вычислительной сложности от вероятности ошибки распознавания подписей, жестов и лиц для решающих алгоритмов на основе иерархического поиска и полного перебора.

Работа поддержана грантом РФФИ № 12-01-00920-а.

- [1] Ланге М. М. Иерархические структуры данных и решающие алгоритмы для классификации изображений // Машинное обучение и анализ данных, 2013. Т.1, №5. — С. 565–573. <http://jmla.org/papers/doc/2013/no5/Lange2013ImageClassification.pdf>.

## Синтез и анализ алгоритма распознавания контурных сигналов с применением комплекснозначных нейронных сетей

*Наумов Александр Сергеевич*<sup>1</sup>

*Роженцов Алексей Аркадьевич*<sup>2</sup>

<sup>1</sup>NaumovAS@volgatech.net, <sup>2</sup>RozhencovAA@volgatech.net

Йошкар-Ола, Поволжский государственный технологический университет

Контур изображения целиком определяет форму изображения и содержит всю необходимую информацию для его распознавания. При этом наиболее адекватной является комплекснозначная модель контуров плоских изображений, сохраняющая и позволяющая извлекать всю необходимую информацию о параметрах линейных преобразований изображений. Поскольку распознаваемые объекты могут обладать сильной вариабельностью формы, для их обработки целесообразно использовать нейронные сети.

Параметрами традиционных нейронных сетей, как правило, являются действительные числа для обработки вещественных сигналов. При обработке плоских изображений групповых точечных объектов и контурных изображений целесообразно представлять их в виде комплекснозначных сигналов. Как результат, в рамках данной работы предлагаются комплекснозначные нейронные сети (КНС), чьи веса (весовые коэффициенты), входные и выходные сигналы являются комплексными числами. В работе рассмотрена модель комплекснозначной нейронной сети (КНС) для распознавания контурных сигналов, заданных в комплексной форме.

Особенностью рассмотренной в работе модели КНС является ее инвариантность к масштабу и повороту входного контурного сигнала. Свойство инвариантности к повороту входного контурного сигнала достигается путем перехода от обработки отсчетов сигнала к последующей обработкой значений состояния нейронов 2-го скрытого слоя, реализуемого при помощи линейной функции активации нейронов 1-го скрытого слоя и Гауссовой функции активации нейронов 2-го скрытого слоя.

Так же в рамках работы рассмотрены вопросы применения алгоритма на базе комплекснозначной нейронной сети в задачах диагностики заболеваний органа зрения по результатам распознавания периметрической картины пациента.

- [1] *Наумов А. С.* Диагностика заболеваний органа зрения по данным периметрических исследований с применением комплекснозначных нейронных сетей // *Нейрокомпьютеры*, М.: РАДИОТЕХНИКА, 2012. — №4. — С. 44–52. <http://www.radiotec.ru/catalog.php?cat=jr7&art=10771>.

## Автоматическая реконструкция трехмерных объектов по ортогональным проекциям

*Захаров Алексей Александрович*<sup>1</sup>

*Жизняков Аркадий Львович*<sup>2</sup>

<sup>1</sup>aa-zaharov@ya.ru, <sup>2</sup>lvovich@newmail.ru

Муром, Муромский институт (филиал) Владимирского государственного университета им. А. Г. и Н. Г. Столетовых

Одной из составляющих электронной модели изделия является компьютерная трехмерная модель объекта, которая может использоваться для разработки управляющих программ, инженерного анализа, визуализации и т. д. Современные САД-системы имеют широкий набор средств для создания трехмерных моделей: булевы операции, операции объектно-ориентированного моделирования, 2.5D-операции, операции модификации вершин, ребер и граней.

Практически все САД-системы позволяют осуществлять генерацию чертежей по трехмерной модели. Однако, получение трехмерной модели по проекциям чертежа вызывает у проектировщиков затруднения, связанные с отсутствием математического и программного обеспечения. Создание системы, выполняющей автоматическую реконструкцию трехмерных моделей по чертежу, позволило бы сократить время проектирования объектов различного назначения [1].

Представляемый алгоритм разработан на основе В-гер-представления, что дает большие возможности по описанию геометрии сложной формы. Основная идея алгоритма состоит в нахождении конструктивных элементов трехмерной модели с помощью заранее определенных шаблонов. Шаблоны описываются при помощи матриц. На основе шаблонов распознаются такие элементы, как отверстия, пазы, фаски и т. д.

Работа поддержана грантом РФФИ № 13-07-00825.

- [1] *Захаров А. А.* Алгоритм автоматической реконструкции трехмерных объектов по ортогональным видам чертежа для САД-систем // Системы управления и информационные технологии, №4(46), 2011. — С. 78–82.

## Метод параметрической реконструкции 3-D плотности распределения облака точек, заданного серией параллельных 2-D сечений

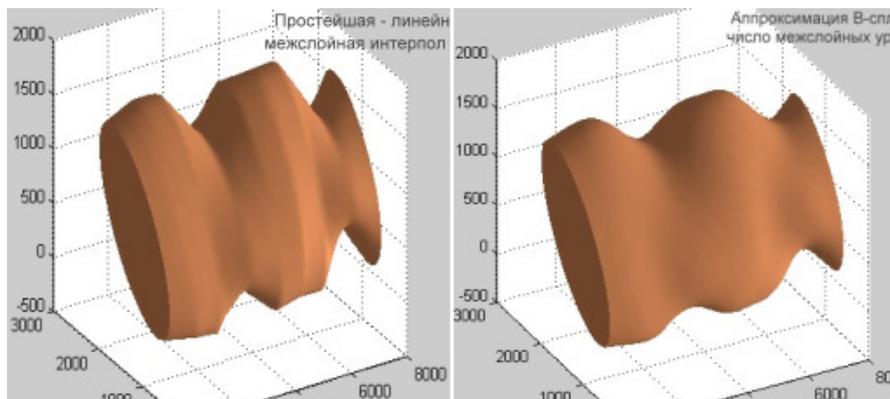
*Анциперов Вячеслав Евгеньевич*<sup>1</sup>

*Евсеев Олег Валерианович*<sup>2</sup>

<sup>1</sup>antciperov@cplire.ru, <sup>2</sup>ev.mipt@gmail.com

Москва, <sup>1</sup>ИРЭ им. В. А. Котельникова РАН, <sup>2</sup>МФТИ

В докладе обсуждаются результаты разработки нового метода реконструкции непрерывных пространственных плотностей распределения мини-объектов (облака точек) по измеренным дискретным данным [1]. Предложенный метод основывается на параметрической межсрезовой интерполяции контуров следа облака в наборе параллельных сечений, рис. 1. Идея метода интерпретируется как задача построения трехмерного распределения плотности вероятностей, согласованной с определенным семейством двумерных условных распределений. Обсуждаются вопросы экономичности метода, связанные с параметрическим характером аппроксимации плотностей 3-D распределений.



**Рис. 1.** Сравнение прямой — межсрезовой линейной интерполяции серии плоских контуров облака точек и предлагаемой параметрической интерполяции B-сплайнами.

Работа поддержана грантом РФФИ № 12-07-00104.

- [1] Анциперов В. Е., Евсеев О. В., Обухов Ю. В. Метод параметрической аппроксимации плотности распределения набора дискретных 3-D вершин, представленных в виде облака точек // Журнал радиоэлектроники, ISSN 1684-1719, М: РАН, ИРЭ им. В. А. Котельникова. — № 8, 2012. <http://jre.cplire.ru/jre/aug12/8/text.pdf>.

## О количественной оценке нестационарности ритмов электроэнцефалограмм в норме и при паркинсонизме в ранних стадиях

*Обухов Юрий Владимирович*<sup>1</sup>

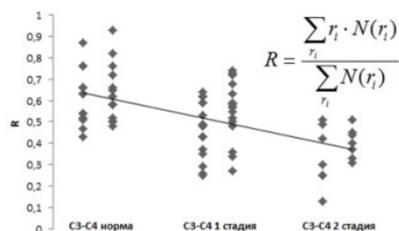
*Королев Михаил Сергеевич*<sup>2</sup>

*Обухов Константин Юрьевич*<sup>3</sup>

<sup>1</sup>obukhov@cplire.ru, <sup>3</sup>ko.arsenalfc@gmail.com

Москва, <sup>1,2</sup>ИРЭ им. В. А. Котельникова РАН, <sup>3</sup>МФТИ

Сигналы ЭЭГ по своей природе являются нестационарными, поэтому представляется целесообразным введение количественной оценки этой нестационарности и сопоставление ее в норме с ранними стадиями БП. Суть предлагаемой оценки заключается в следующем [1]. Вычисляются частотные гистограммы экстремумов пиков в каждом интервале времени длительностью 10 с. Число таких интервалов при ЭЭГ обследовании составляет не менее 12. Оценкой нестационарности является взвешенная сумма коэффициентов корреляции  $r_i$  на их количество  $N(r_i)$  в корреляционной матрице динамических гистограмм экстремумов. На рисунке представлены значения  $R$  в симметричных отведениях ЭЭГ в моторной зоне С3 и С4 для групп практически здоровых людей и пациентов на 1 и 2 стадиях БП по качественной шкале Хен–Яра. Наблюдается снижение среднего по группе значения  $R$ .



Работа поддержана грантом РФФИ № 12-02-00611 и Программой Президиума РАН «Фундаментальные науки – медицине».

- [1] *Обухов Ю. В. и др.* Частотно-временной анализ электрической активности мозга при болезни Паркинсона // В кн. Нейродегенеративные заболевания: фундаментальные и прикладные аспекты под. ред. М. В. Угрюмова — М.: Наука, 2010. — ISBN 978-5-02-036710-4, с. 112–136.

## Совместный анализ частотно-временных характеристик сигналов электроэнцефалографии, электромиографии и механического тремора при болезни Паркинсона в ранних стадиях

Королев Михаил Сергеевич<sup>1</sup>

Обухов Юрий Владимирович<sup>2</sup>

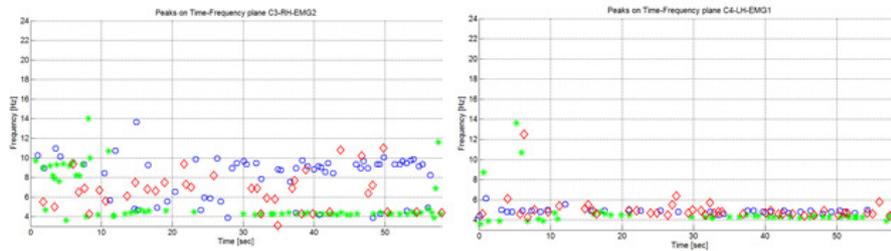
Обухов Константин Юрьевич<sup>3</sup>

Сушкова Ольга Сергеевна<sup>4</sup>

<sup>2</sup>obukhov@splire.ru, <sup>3</sup>ko.arsenalfc@gmail.com

Москва, <sup>1,2,4</sup> ИРЭ им.В.А.Котельникова РАН, <sup>3</sup> МФТИ

Одним из путей поиска признаков болезни Паркинсона (БП) является совместный анализ сигналов разной модальности — электроэнцефалограммы (ЭЭГ), электромиограммы (ЭМГ) и механического тремора (МТ), измеряемого с помощью акселерометров. Частотную синхронизацию ЭЭГ, ЭМГ и МТ можно оценивать по частотно-временному распределению экстремумов вейвлет спектрограмм ЭЭГ [1], МТ и огибающей амплитудно-модулированного высокочастотного ЭМГ.



**Рис. 1.** Локальные максимумы на частотно-временном диапазоне отведений ЭЭГ в моторной зоне коры мозга С3 (кружочки) и контралатеральных МТ (звездочки) и ЭМГ (ромбики) (слева) и меж-полушарно симметричным отведением С4 и контралатеральных МТ и ЭМГ (справа) больного на 1 стадии БП по качественной шкале Хен-Яра.

Из правого рис.1 видно, что экстремумы в больной моторной зоне правого полушария частотно скоррелированы с экстремумами МТ и ЭМГ. Напротив, в еще здоровом левом полушарии такой корреляции нет. Следует отметить, что величина экстремумов МТ в левой руке на два порядка меньше таковой в правой руке.

Работа поддержана грантом РФФИ № 12-02-00611.

[1] *Обухов Ю. В. и др.* Способ ранней электроэнцефалографической диагностики болезни Паркинсона // патент РФ, 20.02.2012, № 2012105777.

## Классификация и распознавание образов в данных магнитной энцефалографии

*Дергузов Аркадий Владимирович*<sup>1</sup>

*Махортых Сергей Александрович*

*Лыжко Екатерина Викторовна*

<sup>1</sup>arkadid@list.ru

Пушино, ИМПБ РАН

В работе предлагаются методы изучения пространственных распределений магнитного поля по поверхности головы с использованием спектральных характеристик в задачах магнитной энцефалографии (МЭГ). Решаются задачи классификации типов активности и диагностики состояния пациента по структуре поля и локализации источников нормальной и патологической биомагнитной активности головного мозга.

В качестве признакового пространства при решении задач классификации типа магнитной активности предлагается использовать функциональный базис, состоящий из сферических гармоник. По ним осуществляется разложение регистрируемых сигналов МЭГ. Выбор информативных признаков осуществляется на этапе обучения алгоритма и основан на оценке изменчивости признаков (коэффициентов Фурье) в классе изучаемых объектов. Для сокращения объема обучающей выборки предлагается ограничить ее только такими распределениями поля, которые не получают друг из друга поворотом. Было предложено несколько схем распознавания записей МЭГ в выбранном признаковом пространстве.

В случаях, имеющих клинические проявления в виде тремора, корректность предлагаемого метода классификации подтверждается независимым анализом миограмм. Полученные результаты и выводы основываются на наличии явных корреляций между записями из предварительно построенной обучающей выборки с априори присутствующей патологией и сигналом в норме с предъявляемой записью. Использовано несколько критериев информативности. Каждый критерий исследовался с точки зрения возможности максимального разделения типов активности. Дальнейший анализ ведется в признаковом пространстве источника, результатом которого является выделение областей значений параметров, характерных для каждого типа активности. Результаты локализации источников патологической активности в записи магнитной энцефалограммы хорошо согласуются с существующим представлением о связи болезни Паркинсона с поражениями подкорковых областей мозга.

Работа выполнена при поддержке РФФИ, гранты №№ 11-07-00519, 11-07-00716, 13-01-00340.

[1] *Derguzov A. V., Makhortykh S. A.* Spectral Analysis and Data Classification in Magnetoencephalography // PRIA, Vol. 16, N. 3, 2006, pp. 497–505.

## Автоматический анализ формы спирографических петель по их сигнатурам

*Манило Людмила Алексеевна*<sup>1</sup>

*Немирко Анатолий Павлович*<sup>2</sup>

*Саламонова Ирина Сергеевна*<sup>3</sup>

<sup>1</sup>lmanilo@yandex.ru, <sup>2</sup>apn-bs@yandex.ru, <sup>3</sup>i.salamonova@yandex.ru

Санкт-Петербург, СПбГЭТУ «ЛЭТИ»

Данное исследование направлено на разработку новых методов и алгоритмов непрерывного контроля функций внешнего дыхания при искусственной вентиляции легких (ИВЛ) по спирографическим петлям, в частности, по петлям «объем-давление» (ОД). Установлено, что этот вид спирограмм представляет наибольший практический интерес для задач мониторинга в процессе проведения респираторной поддержки.

В работе предложены числовые характеристики спирографических петель, позволяющие обнаруживать существенные отклонения в заданных режимах ИВЛ и распознавать ранние формы развития патологий дыхательных путей. Кроме того, показана возможность косвенной оценки изменений динамической растяжимости и сопротивления дыхательных путей по числовым параметрам сигнатуры петель. Создана автоматизированная база спирографических данных, включающая сигналы, полученные на аппарате ИВЛ путем механического моделирования изменений растяжимости  $C$  и сопротивления  $R$  дыхательных путей, а также реальные записи спирограмм, зарегистрированные в условиях клиники. Проведены эксперименты по оценке эффективности распознавания различных форм петель по сигнатуре (реальные записи спирограмм) и обнаружению значимых изменений параметров вентиляции  $C$  и  $R$  (модельные сигналы).

Разработан комплекс программ автоматического анализа спирограмм, который предполагается использовать в аппаратах ИВЛ для расширения их диагностических возможностей. Как показали исследования в клинических условиях, постоянный контроль за формой кривой ОД, ее наклоном и шириной являются важной составляющей респираторной поддержки, особенно у пациентов с развивающейся патологией легких.

Работа поддержана государственным контрактом Министерства образования и науки РФ № 16.522.12.2016 и грантами РФФИ № 12-01-00583, № 13-01-00540.

- [1] *Манило Л. А.* Автоматический анализ формы спирографических петель по их сигнатурам // Машинное обучение и анализ данных, 2013.

## О комбинаторном тестировании условия локальной полноты связных размеченных графов в применении к задачам хемоинформатики

*Торшин Иван Юрьевич*<sup>1</sup>

*Рудаков Константин Владимирович*<sup>2</sup>

<sup>1</sup>tiy135@yahoo.ca, <sup>2</sup>rudakov@ccas.ru

Москва, <sup>1,2</sup>МФТИ, <sup>2</sup>ВЦ РАН

Комбинаторная теория разрешимости, представляющая собой развитие алгебраического подхода к задачам распознавания, является современным инструментом для исследования признаков описаний объектов. В докладе будет представлен формализм, позволяющий применять комбинаторную теорию разрешимости к теоретико-графовым построениям в области хемоинформатики.

Показано, что в свете фундаментальных физико-химических особенностей строения молекул целесообразно ввести специальное понятие  $\chi$ -графа (хемографа) — особой разновидности размеченного графа для описания связности молекул. Рассмотрены фундаментальные свойства хемографов, введены особые виды разметок  $\chi$ -подграфов —  $\chi$ -цепи (цепи размеченных вершин) и  $\chi$ -узлы (подграфы окрестностей размеченных вершин). На основе основных постулатов теории химической связи проведен анализ аксиоматики для введения функций разметки хемографов. Показано, что отношения вхождения  $\chi$ -цепей и  $\chi$ -узлов в хемограф являются инвариантами хемографов; получены критерии полноты соответствующих инвариантов. В рамках комбинаторной теории разрешимости  $\chi$ -графы рассматриваются как объекты, а их инварианты — как признаковые описания объектов. Получены критерии локальной полноты исследуемых наборов инвариантов  $\chi$ -графов (т.е. полноты относительно заданного множества прецедентов). Показано, что комбинаторное тестирование критерия регулярности соответствующей задачи распознавания позволяет количественно оценить локальную полноту используемых  $\chi$ -инвариантов. Представлены результаты практического применения предлагаемого формализма к одной из задач хемоинформатики — поиску химических соединений, структурно схожих с заданным.

Работа поддержана грантом РФФИ № 12-07-00485.

- [1] *Торшин И. Ю., Рудаков К. В.* О применении комбинаторной теории разрешимости к задачам анализа хемографов. Основы современной теории химической связи и понятие хемографа // Pattern Recognition and Image Analysis, М.: МАИК Наука. — 2014 (принято к печати).

## Двухфазная схема распознавания в задаче «структура–свойство»

*Прохоров Евгений Игоревич*

*Кумсков Михаил Иванович*

qsar\_msu@mail.ru

Москва, Московский государственный университет имени М. В. Ломоносова, механико-математический факультет

Работа посвящена методам поиска количественных соотношений «структура–свойство» для прогнозирования активности химических соединений. Ключевой особенностью задачи является ее ориентированность на предсказание активности новых соединений. В связи с этим авторами предложено использование правил отказа для распознающих моделей (ограничений допустимости).

Двухфазная схема решения заключается в следующем. Пусть построена распознающая модель, решающая исходную задачу классификации (модель первого уровня) и вычислен её функционал качества  $\varphi_1$  (отношение количество верно спрогнозированных соединений к общему числу осуществленных прогнозов). Сформулируем задачу классификации второго уровня. В первый класс войдут те соединения, прогноз активности которых моделью первого уровня осуществлен верно, а во второй класс – неверно. Пусть для задачи второго уровня также построена распознающая модель и её функционал качества равен  $\varphi_2$ .

Определим теперь результирующую модель. Результирующая модель осуществляет отказ от прогноза данного соединения в том случае, если модель второго уровня относит данное соединение ко второму классу и осуществляет прогноз активности моделью первого уровня в противном случае. Обозначим функционал качества результирующей модели через  $\varphi_0$ , а через Reject – количество отказов от прогноза. Тогда верна следующая оценка качества результирующей модели.

$$\varphi_0 = \frac{(\varphi_1 + \varphi_2)N - \text{Reject}}{2(N - \text{Reject})}.$$

В частности  $\varphi_0 > \varphi_1$ , если  $\varphi_2 \geq \varphi_1 > 1/2$  и  $\text{Reject} > 0$  или если  $\varphi_2 > \varphi_1 > 1/2$ . Таким образом, доказано улучшение качества прогноза при использовании нетривиальных правил отказа от прогноза. Также в ходе работы проведены многочисленные испытания, подтверждающие практическую эффективность разработанного подхода.

Работа поддержана грантом РФФИ № 10-07-00694.

- [1] Прохоров Е.И. Оценка качества прогноза для двухфазной схемы решения задачи «структура–свойство» // Вестник Московского университета. Серия 1: Математика. Механика. 2013. Поступила в редакцию 11.02.2013.

## Спектральный подход к распознаванию повторов в геномах и белках

*Панкратов Антон Николаевич*<sup>1</sup>

*Пятков Максим Иванович*

*Руднев Владимир Ремович*

*Куликова Людмила Ивановна*

<sup>1</sup>pan@impb.ru

Пушчино, Институт математических проблем биологии РАН

Разрабатывается спектральный подход к определению повторов в биоинформационных данных и получению на его основе новых знаний о повторах. Рассматриваемый с общих позиций, подход заключается в оценке сходства некоторых характеристик вдоль цепи макромолекулы. Само сходство оценивается интегрально как интеграл от квадрата отклонения функции-признака. При этом ключевым параметром становится размер окна, в котором оценивается сходство. При минимальном размере окна сходство переходит в точечную оценку, когда сравниваются свойства отдельных оснований полимерной молекулы. Оптимальное для каждой задачи значение окна позволяет избежать флуктуаций, присущих точечным оценкам. Соответственно этому предложению возникает понятие матрицы сходства как аналога хорошо известного способа отображения гомологии на точечной матрице сходства.

Оценивание интеграла квадрата отклонения производится на основе разложения исследуемых функций-признаков в ряд Фурье. За счет этого достигается, с одной стороны, высокая вычислительная эффективность предложенного метода, поскольку для оценки интеграла во многих задачах достаточно несколько коэффициентов разложения, а с другой стороны, дополнительное сглаживание функций-признаков.

В докладе рассмотрены примеры применения данного математического аппарата к задачам распознавания различных типов повторов в геномах, распознаванию супервторичных структур белковых молекул, обладающих уникальной пространственной укладкой, а также прогнозированию возникновения таких структур по аминокислотной последовательности белков.

Работа выполняется при поддержке гранта РФФИ № 11-07-00716.

- [1] *Pankratov A. N., Gorchakov M. A., Dedus F. F., Dolotova N. S., Kulikova L. I., Makhortykh S. A., Nazipova N. N., Novikova D. A., Olshevets M. M., Pyatkov M. I., Rudnev V. R., Tetuev R. K., Filippov V. V.* Spectral Analysis for Identification and Visualization of Repeats in Genetic Sequences // Pattern Recognition and Image Analysis, 2009, vol. 19, № 4, pp. 687–692.

## Распознавание и анализ устойчивости структурных мотивов типа $\alpha$ - $\alpha$ -уголок в глобулярных белках

*Панкратов Антон Николаевич*<sup>1</sup>

*Руднев Владимир Ремович*<sup>2</sup>

*Куликова Людмила Ивановна*<sup>3</sup>

*Дедус Флоренц Федорович*<sup>4</sup>

*Тихонов Дмитрий Анатольевич*<sup>5</sup>

*Ефимов Александр Васильевич*<sup>6</sup>

<sup>1</sup>pan@impb.ru

Пушино, <sup>1,2,3,4,5</sup>Институт математических проблем биологии РАН,

<sup>6</sup>Институт белка РАН

Важной теоретической и прикладной задачей молекулярной биологии является изучение и предсказание элементов пространственной структуры белковых молекул, в частности, распознавание супервторичных элементов пространственной структуры белковой глобулы. В данной работе рассматривается задача распознавания структурных мотивов белков на примере  $\alpha$ - $\alpha$ -уголков. Эта супервторичная структура образована двумя соседними по полипептидной цепи  $\alpha$ -спиралями, связанными между собой перетяжкой и упакованных ортогонально в пространстве.

Разработан комбинированный подход к решению задач анализа пространственной структуры белков на основе аналитического описания основной цепи белковой глобулы и спектрального метода распознавания повторов. Найдено 110  $\alpha$ - $\alpha$ -уголков в базе данных PDB, соответствующих заданному эталону. Методом молекулярной динамики показано, что  $\alpha$ - $\alpha$ -уголки как автономные структуры устойчивы в водной среде.

Из автономной устойчивости следует справедливость утверждений:

- все признаки структуры локализованы в соответствующем участке первичной структуры белковой молекулы;
- свойство устойчивости можно использовать для верификации этой структуры;

что свидетельствует о хорошей обусловленности задачи распознавания данной структуры.

Работа выполняется при поддержке грантов РФФИ № 11-07-00716, 13-01-00340, 11-07-00519.

- [1] Руднев В. Р., Панкратов А. Н., Куликова Л. И., Дедус Ф. Ф., Тихонов Д. А., Ефимов А. В. Распознавание и анализ устойчивости структурных мотивов типа  $\alpha$  -  $\alpha$ -уголок в глобулярных белках // Математическая биология и биоинформатика, 2013. —Т. 8. — № 2. — С. 398–406. [http://www.matbio.org/2013/Rudnev\\_8\\_398.pdf](http://www.matbio.org/2013/Rudnev_8_398.pdf).

## **Анализ структуры зависимостей в корпусе взаимосвязанных текстов с применением системы «Антиплагиат» и интеллектуального анализа данных на примере диссертаций в области исторических наук**

*Ботов П. В.<sup>1</sup>, Вьючков Д. В.<sup>2</sup>, Корольков М. Е.<sup>3</sup>,  
Песков Н. В.<sup>4</sup>, Сурувенко Н. С.<sup>5</sup>, Хританков А. С.<sup>6</sup>,  
Царьков С. В.<sup>7</sup>, Чехович Ю. В.<sup>8</sup>*

<sup>6</sup>anton.khritanikov@acm.org

Москва, <sup>1,2,4,5,6,7,8</sup>ЗАО «Анти-Плагиат», <sup>3</sup>ВМК МГУ, <sup>6,7,8</sup>МФТИ, <sup>8</sup>ВЦ РАН

Объектом исследования являются диссертации в области исторических наук из базы диссертаций Российской государственной библиотеки. Анализировались сами тексты, их метаданные, а также взаимосвязи текстов — заимствования, цитирование тех же источников, цитирование других работ внутри базы.

Целью работы было выявление статистических закономерностей и кластерных структур в заимствованиях между текстами диссертаций, защищенных по историческим наукам.

Для подготовки данных использована система «Антиплагиат». Выполнен поиск всех заимствований с длиной фрагмента не менее 4 слов и не менее 30 символов. Близко стоящие фрагменты объединялись, фрагменты длиной менее 250 символов отбрасывались. В дальнейшем анализе не учитывались титульная страница и список литературы по их расположению в текстах.

Корректно оформленные цитирования были исключены в два этапа: сначала эвристически выделялись оформленные по правилам русского языка блоки текста, для определения цитат к ним применялся алгоритм классификации бинарного дерева. Достигнуты следующие показатели качества выделения цитат: точность 95.8%, полнота 43.8%.

Анализ диссертаций показал, что большая часть работ (90%) не имеет значимых заимствований (менее 10%). Только 2.25% диссертаций имеют больше 50% заимствований, и лишь 1% работ попадает в диапазон от 70% до 100%.

К графу попарных заимствований между текстами были применены методы анализа сетевых структур (SNA). С помощью инструмента Gephi были выделены и визуализированы слабо связанные компоненты и сообщества. Параметры алгоритмов выбраны так, чтобы количество компонент было максимальным. Размер наибольшего сообщества составил около 2000 узлов, всего выделено 748 сообществ. Большое внимание уделялось наглядной интерпретации полученных кластеров.

## Аддитивная регуляризация тематических моделей

*Воронцов Константин Вячеславович*

voron@forecsys.ru

Москва, ВЦ РАН, МФТИ, НИУ ВШЭ, ВМК МГУ

Тематическое моделирование — одно из активно развивающихся направлений статистического анализа текстов. *Вероятностная тематическая модель* выявляет тематику коллекции текстовых документов, описывая каждую тему дискретным распределением на множестве терминов, каждый документ — дискретным распределением на множестве тем.

Модель латентного размещения Дирихле LDA (Блэй и др., 2003) служит основой для сотен специализированных модификаций, но порождает две открытые проблемы, редко упоминаемые в литературе. Во-первых, априорное распределение Дирихле имеет крайне слабые лингвистические обоснования. Популярность модели LDA объясняется скорее чисто математическим удобством аналитического интегрирования в байесовском выводе. Во-вторых, для практических задач необходимы композитные модели, совмещающие большое число функциональных требований. В частности, для научного поиска по большим коллекциям публикаций нужна модель, одновременно иерархическая, динамическая,  $n$ -граммная, разреженная, робастная, мультиязычная, частично обучаемая, учитывающая авторство и цитирование, и т. д. Байесовский вывод оказывается слишком громоздким для совмещения более 2–3 требований в одной модели. Такие модели в литературе пока не рассматривались. Назрела необходимость разработки новых принципов построения тематических моделей, свободных от избыточных вероятностных допущений.

Предлагаемая теория *аддитивной регуляризации тематических моделей* (АРТМ) решает эти проблемы. Задача тематического моделирования является некорректно поставленной. Для её решения в АРТМ предлагается применять не байесовскую, а более общую тихоновскую регуляризацию. Каждое требование формализуется в виде штрафного слагаемого к функционалу логарифма правдоподобия и приводит к аддитивной поправке в формулах М-шага EM-алгоритма. Распределение Дирихле утрачивает центральную роль. Это лишь один из возможных регуляризаторов, не самый лучший и не настолько универсальный, как принято считать. В докладе рассматриваются различные проблемно-ориентированные регуляризаторы и показывается, что АРТМ позволяет намного проще выводить многие известные модели, а также строить гораздо более сложные многофункциональные и композитные модели.

Работа выполнена при поддержке РФФИ, проект №11-07-00480.

- [1] *Воронцов К. В.* Аддитивная регуляризация тематических моделей // Доклады РАН. — 2013 (в печати).

## Разреживание вероятностных тематических моделей

*Потапенко Анна Александровна*

anya\_potapenko@mail.ru

Москва, ВМК МГУ

Вероятностная тематическая модель (ВТМ) коллекции текстовых документов описывает каждую тему дискретным распределением на множестве терминов, каждый документ — дискретным распределением на множестве тем. Естественно полагать, что каждый документ относится к небольшому числу тем, и каждая тема определяется небольшой долей характерных для неё терминов. Разреженные модели позволяют строить сжатые представления больших коллекций, эффективнее решать задачи информационного поиска, классификации и категоризации текстов.

Большинство известных подходов к разреживанию основаны на модели LDA, что приводит к искусственному конфликту между сглаживанием и разреживанием, и усложнению модели. Предлагается простой метод постепенного обнуления наименьших вероятностей терминов в темах и тем в документах в ходе итераций EM-алгоритма. Ближайшим аналогом является метод разреживания нейронных сетей OBD (optimal brain damage). Другим обоснованием разреживания является введение регуляризатора, который поощряет распределения, максимально непохожие на равномерное, то есть имеющие минимальную энтропию.

При сильном разреживании тематическая модель не способна описывать некоторые слова в документах. Эта проблема решается с помощью робастных моделей, которые вводят шумовые и фоновые слова наряду с тематическими. Рассматриваются две робастные модели: в первой распределения шума и фона оптимизируются; во второй (упрощённой) модели каждому слову в документе жёстко назначается роль либо шумового, либо тематического. Эксперименты показывают, что разреженная робастная модель лучше классической сглаженной модели LDA, причём упрощённой модели вполне достаточно.

В докладе приводятся результаты экспериментов на двух коллекциях научных публикаций, русскоязычной и англоязычной. Исследуется несколько стратегий разреживания, позволяющих достигать разреженности до 99% при 100 темах, улучшая качество модели на контрольной выборке. Столь высокая доля нулевых вероятностей терминов в темах означает, что каждый термин относится в среднем к одной теме, причём многие термины автоматически отбрасываются как нетематические.

Работа выполнена при поддержке РФФИ, проект №11-07-00480.

- [1] *Воронцов К. В., Потапенко А. А.* Модификации EM-алгоритма для вероятностного тематического моделирования // Машинное обучение и анализ данных (<http://jmla.org>). — 2013.

## Методы вычисления релевантности фрагментов текста на основе тематических моделей в задаче автоматического аннотирования

*Царёв Дмитрий Владимирович*<sup>1</sup>

*Машечкин Игорь Валерьевич*<sup>2</sup>

*Петровский Михаил Игоревич*<sup>3</sup>

<sup>1</sup>tsarev@mlab.cs.msu.su, <sup>2</sup>mash@cs.msu.su, <sup>3</sup>michael@cs.msu.su

Москва, Факультет вычислительной математики и кибернетики Московского государственного университета имени М.В.Ломоносова

В статье рассматриваются наиболее актуальные методы вычисления релевантности (значимости) фрагментов текста на основе анализа тематических моделей для последующего построения аннотаций в форме выдержек, т. е. аннотаций, полностью состоящих из последовательности фрагментов исходного текста. Кроме того, представлен собственный разработанный метод вычисления релевантности фрагментов текста, основанный на оценке весов тематик в нормализованном пространстве тематик, получаемом с помощью неотрицательной матричной факторизации, которая используется в качестве матричного разложения в модели латентно-семантического анализа. Эксперименты, проведенные с методами автоматического аннотирования на эталонных тестовых наборах DUC 2001, DUC 2002 с использованием стандартных метрик оценки качества аннотаций ROUGE, показали превосходство предложенного метода по качеству получаемых аннотаций.

Используя релевантность фрагментов текста в качестве оценок количества информации, содержащейся в них, были проведены экспериментальные исследования по предварительному информационному сжатию текстов (выбору наиболее значимых предложений) и последующей их многотемной классификации и кластеризации. Эксперименты показали улучшение качества классификации и кластеризации на эталонных тестовых наборах Reuters-21578 (для многотемной классификации) и 20 Newsgroups (для кластеризации). Исходя из полученных результатов, можно сделать вывод о применимости предложенного подхода предварительного информационного сжатия текстов в рассмотренных задачах интеллектуального анализа текстовых данных (англ. text mining).

Работа выполнена при поддержке государственного контракта № 14.514.11.4016 и грантов РФФИ 11-07-00616, 12-07-00585.

- [1] *Машечкин И. В., Петровский М. И., Царёв Д. В.* Методы вычисления релевантности фрагментов текста на основе тематических моделей в задаче автоматического аннотирования // Вычислительные методы и программирование. Новые вычислительные технологии. — Т. 14, 2013. — С. 91–102. [http://num-meth.srcc.msu.ru/zhurnal/tom\\_2013/pdf/v14r112.pdf](http://num-meth.srcc.msu.ru/zhurnal/tom_2013/pdf/v14r112.pdf).

## Извлечение тегов веб-страниц с частичным обучением

*Лексин Василий Алексеевич*<sup>1</sup>

*Николенко Сергей Игоревич*<sup>2</sup>

<sup>1</sup>vasily.leksin@gmail.com, <sup>2</sup>sergey@logic.pdmi.ras.ru

Москва, ООО «Сёрфингберд»

Особенностью рекомендательного сервиса веб-сёрфинга (Surfingbird, StumbleUpon) от традиционных рекомендательных систем с медийным контентом (Netflix, Last.fm и др.) является обилие текстового контента. Методы анализа текста позволяют эффективно решать проблему холодного старта, присущую всем традиционным методам коллаборативной фильтрации, которые строят рекомендации только на основе рейтингов пользователей без анализа содержимого объектов. Одним из важных аспектов методов текстового анализа является автоматическое выявление тегов (ключевых фраз) для веб-страниц, попадающих в рекомендательный сервис. Теги позволяют легко находить схожие объекты, рекомендовать пользователю самые свежие веб-страницы, соответствующие его интересам, а также строить облако тегов пользователя.

Традиционно, задачу рекомендаций тегов принято рассматривать для случая, когда для большинства веб-страниц теги уже расставлены самими пользователями в момент добавления страницы в систему. Так происходит, например, в системах социальных закладок, где задача заключается лишь в том, чтобы найти недостающие теги, исправить ошибки правописания и так далее. В нашем случае теги для большинства веб-страниц неизвестны: новые страницы чаще всего попадают в базу автоматически, а пользователи, прочитавшие веб-страницу вряд ли будут тратить время на заполнение тегов для нее. Таким образом, мы решаем более сложную задачу: рекомендовать теги веб-страниц на основе анализа текста и используя обучающую выборку для небольшой доли размеченных тегами веб-сайтов (преимущественно, это страницы, попадающие через RSS-потоки). Для решения этой задачи мы используем методы классификации с частичным обучением. Для обучения используются два набора данных: точно размеченные веб-страницы (страницы из RSS-потоков с известными тегами) и частично размеченные веб-страницы, теги для которых получены автоматически из их контента. Также мы произвели сравнение нескольких алгоритмов классификации с частичным обучением. Результаты экспериментов показали, что используемый метод показывает хорошие результаты (для оценки качества считалась  $F$ -мера на контрольной выборке) даже в случае малой доли размеченных веб-страниц (в нашем случае 5–10% от их общего числа).

[1] *Leksin V. A., Nikolenko S. I.* Semi-supervised tag extraction in a web recommender system // 6-th Int'l. Conf. SISAP-2013. — Pp. 206–212.

## Байесовский подход к оцениванию факторов, влияющих на положение сайта в результатах поискового запроса

*Нгуен Тронг Тинь*<sup>1</sup>

*Поленова Елена Александровна*<sup>2</sup>

*Красоткина Ольга Вячеславовна*<sup>3</sup>

*Моттль Вадим Вячеславович*<sup>4</sup>

<sup>1</sup>nguyentrongtinh7512@yahoo.com.vn, <sup>3</sup>o.v.krasotkina@yandex.ru,

<sup>4</sup>vmottl@yandex.ru

Тула, <sup>1,2,3,4</sup>Тульский государственный университет; Москва, <sup>3,4</sup>ВЦ РАН

Сегодня основным инструментом доступа к информации в интернете являются поисковые системы, которые, принимая запрос пользователя, возвращают упорядоченный список web-страниц. Каждая поисковая система имеет свой собственный очень сложный и постоянно совершенствующийся алгоритм ранжирования, который является ее коммерческой тайной. Очевидно, что чем выше находится сайт в рейтинге поисковой системы по определенным запросам, тем больше посетителей будет у ресурса. Специалисты по продвижению сайтов в процессе работы сталкиваются с тремя основными проблемами. Во-первых, современные алгоритмы ранжирования зависят не только от наполнения веб-страниц, но и от поискового запроса, поэтому зачастую методика продвижения, сработавшая с одним ресурсом, может не дать результата с другим. Во-вторых, до момента очередной индексации специалист не может видеть, как его усилия по продвижению повлияли на позицию сайта. В-третьих, среди огромного количества признаков, характеризующих пару запрос-документ, необходимо выбрать такие, изменение которых способно повлиять на позицию сайта максимальным образом.

Целью данной работы является разработка математического аппарата, который позволит создать инструмент, позволяющий восстанавливать формулу ранжирования поисковой системы для конкретного поискового запроса и определять факторы, наиболее влияющие на положение сайта в поисковой выдаче. В данной работе предлагается байесовская концепция восстановления регрессионной зависимости в случае, когда выходная переменная представлена в порядковой шкале. Основная идея обучения заключается в использовании параметрического семейства априорных распределений объектов в линейном пространстве признаков вместе с априорными распределениями параметров решающей функции, что приводит к эффекту селективности признаков.

- [1] *Красоткина О.В., Нгуен Т. Ч., Поленова Е. А., Моттль В. В.* Байесовский подход к оцениванию факторов, влияющих на положение сайта в результатах поискового запроса // Известия Тульского государственного университета. Технические науки, выпуск 2, 2013, С. 168–177.

## Частичный синтаксический разбор с помощью условных случайных полей

*Кудинов Михаил Сергеевич*

mikhailkudinov@gmail.com

Москва, Вычислительный центр им. А. А. Дородницына РАН, Управление высокопроизводительных алгоритмов Исследовательского центра «Самсунг»

В статье изложен подход к поиску синтаксически связанных групп соседних слов (chunks) в русском тексте. Продемонстрирована принципиальная возможность и корректность постановки задачи выделения таких групп применительно к языку со свободным порядком слов. С использованием аппарата условных случайных полей определенный класс подобных групп можно выделить с  $F_1$ -мерой не менее 0.94. При этом обучающая выборка может быть получена путем обработки исходного текста синтаксическим анализатором без последующей ручной коррекции результатов. Тем не менее, выделение достаточно длинных фрагментов текста оказывается затруднительным, а показатель  $F_1$ -меры, полученный в эксперименте, достаточно низким.

Таким образом, результаты показывают, что метод работает достаточно надежно и может найти свое применение в тех задачах, где применение полного синтаксического анализа не требуется. Этими задачами могут быть поиск ключевых слов в документе или высказывании, поступающем на вход диалоговой системы. Метод позволяет выделять в тексте такие фрагменты, как: [экологическая программа], [Президиум Совета Министров СССР], [бассейн Азовского моря].

Достоинством метода является его нетребовательность к качеству обучающей выборки: она может быть получена на основе выдачи доступного синтаксического анализатора. Недостатком является низкая точность выделения фрагментов, содержащих внутри предлог и/или союз: [Комиссия ООН] по [правам человека] вместо [Комиссия ООН по правам человека]. В целом можно констатировать, что при внесении необходимых поправок, зависящих от языка, метод работает не хуже, чем для языков с более строгим порядком слов.

- [1] *Кудинов М. С.* Частичный синтаксический разбор текста на русском языке с помощью условных случайных полей // Машинное обучение и анализ данных, 2013. <http://jmla.org>

## Использование правил со сложной структурой для коррекции документов в формате LaTeX

*Чувилин Кирилл Владимирович*

kirill.chuvilin@gmail.com

Москва, Московский физико-технический институт (ГУ)

Многие конференции и издательства принимают материалы от авторов в формате LaTeX. В каждом издательстве есть определенные традиции и требования к оформлению элементов публикуемого материала: заголовков, списков, таблиц, библиографии, формул, чисел, и т.п. Ошибки, связанные с несоблюдением этих правил, называются *типографическими*. При текущем уровне технологий исправление таких ошибок производится корректорами вручную. Предлагаемый подход направлен на значительное сокращение объема рутинной работы. Для этого каждое правило формализуется. Оказывается, что вручную описывать структуру правил неэффективно, поэтому правила строятся автоматически, основываясь на уже обработанных документах. Отображения вершин синтаксических деревьев документов, не прошедших корректорскую правку, в вершины деревьев обработанных документов составляют обучающую выборку, по которой синтезируются правила.

В предыдущих работах описывался синтез правил, обладающих *простой структурой*: каждое правило реализует операцию удаления, добавления или изменения одной вершины синтаксического дерева и использует линейные последовательности вершин для выбора позиции применения. Несмотря на то, что построенный набор правил позволял автоматически обнаруживать в среднем одну треть часть всех ошибок, оценки его точности и полноты не превосходили 50%.

В данной работе рассматривается два подхода к усложнению структуры правил, которые позволяют значительно повысить точность и полноту синтезируемого набора. Первый подход связан с построением *групповых правил*, как совокупности двух существующих, имеющих близкие позиции применимости. В рамках второго подхода рассматриваются нелинейные шаблоны для поиска мест применимости правил: каждому правилу сопоставляются два *шаблонных дерева*, которые описывают требуемую структуру синтаксического дерева документа рядом с тем узлом, к которому правило может быть применено. Оценки качества набора правил строятся при последовательном наращивании обучающей выборки.

- [1] Чувилин К. В. Использование правил со сложной структурой для коррекции документов в формате LaTeX // Машинное обучение и анализ данных, Город: Изд-во, 2013. — С. 5–25. <http://jmla.org/papers/doc/2013/no6>.

## Вейвлет-технология обработки и анализа геомагнитных данных

*Мандрикова Оксана Викторовна*<sup>1,2</sup>

*Соловьев Игорь Сергеевич*<sup>1,2</sup>

oksanam1@mail.ru, kamigsol@yandex.ru

Петропавловск-Камчатский, <sup>1</sup>Институт космических исследований и распространения радиоволн ДВО РАН, <sup>2</sup>Камчатский государственный технический университет

Работа направлена на создание программных систем по анализу геомагнитных данных и выделению возмущений, обусловленных повышенной активностью Солнца. Воздействие солнечного ветра на магнитосферу Земли приводит к изменению параметров магнитного поля и вызывает возбуждение разного рода волн. В эти моменты времени в регистрируемых геомагнитных данных возникают негладкие локальные структуры, характеризующие степень возмущенности магнитного поля. Сложная структура данных делает неэффективными для их изучения традиционные методы анализа временных рядов, не позволяющие получать информацию о локальных изменениях, протекающих в физическом процессе. Авторами предложена технология анализа геомагнитных данных, основанная на вейвлет-преобразовании, и позволяющая в автоматическом режиме выделять локальные структуры, возникающие в периоды возмущений поля, и оценивать энергетические характеристики магнитного поля [1]. Процедура выделения локальных структур основана на конструкции вейвлет-пакетов и применении пороговых функций. Выделение периодов повышенной геомагнитной активности и оценка интенсивности возмущений поля выполняется на основе непрерывного вейвлет-преобразования.

Апробация технологии выполнена на данных магнитного поля Земли, полученных в обсерватории «Паратунка» (с. Паратунка, Камчатский край). В регистрируемых вариациях геомагнитного поля выделены частотно-временные интервалы, содержащие слабые и сильные геомагнитные возмущения и выполнена оценка интенсивности возмущений. Во время развития магнитной бури интенсивность возмущений значительно увеличивается, что позволяет в автоматическом режиме фиксировать момент предстоящей бури.

Работа поддержана грантом РФФИ № 11-07-98514-р\_восток\_а.

- [1] Мандрикова О. В., Соловьев И. С. Вейвлет-технология обработки и анализа геомагнитных данных // Цифровая обработка сигналов — 2012. №2. — С. 24–29.

## Моделирование и анализ параметров ионосферы на основе совмещения вейвлет-преобразования и авторегрессионных моделей

*Мандрикова Оксана Викторовна*<sup>1</sup>

*Глушкова Надежда Владимировна*<sup>2</sup>

*Полозов Юрий Александрович*<sup>3</sup>

<sup>1</sup>oksanam1@mail.ru, <sup>2</sup>nv.glushkova@yandex.ru, <sup>3</sup>up\_agent@mail.ru

Петропавловск-Камчатский, Институт космических исследований и распространения радиоволн ДВО РАН, Камчатский государственный технический университет

Работа направлена на создание средств анализа ионосферных параметров и выявления аномалий, возникающих в периоды ионосферных возмущений. Структура ионосферы является изменчивой и неоднородной, её изучение основано на анализе вариаций регистрируемых параметров среды. Сложный характер исследуемых процессов, их априорная неопределенность и, как следствие, сложная структура регистрируемых данных требует наличия целого комплекса методов и технологий, позволяющих выполнять моделирование, структурный анализ данных и интерпретацию получаемых результатов.

В работе предложен метод моделирования и анализа ионосферных параметров, основанный на совмещении вейвлет-преобразования с моделями авторегрессии — проинтегрированного скользящего среднего [1]. Метод позволяет выявлять закономерности в параметрах ионосферы и получать прогноз о вариациях. На основе конструкции дискретного вейвлет-преобразования и применения пороговых функций, разработан алгоритм, позволяющий в автоматическом режиме выделять аномалии в ионосфере, оценивать их длительность, интенсивность и масштабы [1].

Апробация предложенных средств выполнялась на данных критической частоты ионосферы  $f_0F_2$ , регистрируемых на Камчатке и в Магадане. Построенные модели естественного хода параметров ионосферы позволили выполнить анализ её динамического режима и построить прогноз с шагом упреждения до пяти часов. На основе оценки отклонения от фонового уровня выявлены аномалии в ионосфере длительностью от нескольких десятков минут до нескольких часов, возникающие в периоды повышенной солнечной активности, а также накануне и в моменты сильных землетрясений на Камчатке.

Работа поддержана грантом РФФИ № 11-07-98514-р\_восток\_а.

- [1] Мандрикова О. В., Глушкова Н. В., Полозов Ю. А. Алгоритмы выделения и анализа аномалий в параметрах критической частоты ионосферы  $f_0F_2$  на основе совмещения вейвлет-преобразования и авторегрессионных моделей // Цифровая обработка сигналов, 2013. №1 — С. 47–53.

## Построение вычислительной процедуры комплексной проверки подлинности данных ДЗЗ

*Кузнецов Андрей Владимирович*<sup>1</sup>

*Мясников Владислав Валерьевич*<sup>2</sup>

<sup>1</sup>kuznetsov@outlook.com, <sup>2</sup>vmyas@smr.ru

Самара, Институт систем обработки изображений РАН

Данные, получаемые с космических аппаратов (КА) дистанционного зондирования Земли (ДЗЗ), включают в себя две составляющие: цифровое изображение и соответствующие этому изображению метаданные. Изменениям после получения может подвергаться как собственно изображение, так и метаданные. Под проверкой подлинности данных ДЗЗ далее будем понимать комплекс мероприятий, позволяющих ответить на вопрос: «Были ли внесены изменения в данные ДЗЗ?». В настоящее время существуют два основных подхода к проверке подлинности цифровых изображений в общем случае и данных ДЗЗ в частности: активный и пассивный.

В первом разделе работы описана формальная постановка задачи пассивной защиты данных ДЗЗ – вводятся основные определения и элементарные алгоритмы проверки подлинности. Второй раздел посвящён описанию вычислительной процедуры комплексной проверки подлинности данных ДЗЗ и показателей критерия оптимальности. В третьем разделе приводится описание точного и приближённого алгоритмов построения вычислительной процедуры при наличии данных о статистике срабатываний ЭА-ов. В этом разделе также описано сравнение разработанных алгоритмов и приведены результаты экспериментов. Четвёртый раздел посвящён описанию алгоритма построения вычислительной процедуры при отсутствии данных о статистике срабатываний ЭА-ов. В пятом разделе приводится описание построения вычислительной процедуры распознавания атаки.

Работа поддержана грантом РФФИ № 12-07-00021-а, программой фундаментальных исследований Президиума РАН «Фундаментальные проблемы информатики и информационных технологий» № 2.12, Министерством образования и науки РФ (в рамках постановления Правительства РФ от 09.04.2010 г. № 218).

- [1] *Кузнецов А. В., Мясников В. В.* Построение вычислительной процедуры комплексной проверки подлинности данных ДЗЗ // Компьютерная оптика, № 37(2), 2013. — С. 245–254. (в печати)

## Использование искусственных нейроподобных сетей для управления технологическими процессами железо-обоганительного комплекса

*Зарубин Михаил Юрьевич, Зарубина Венера Равиловна*  
zarubin\_mu@mail.ru, zarubina\_v@mail.ru

Казахстан, г. Рудный, Рудненский индустриальный институт

Для технических систем основным показателем эффективности внедрения той или иной схемы автоматизации является снижение себестоимости продукции. Использование аппарата ИНС в таких системах способно решить проблему выявления сложных закономерностей и выбора значимых управляющих факторов. Обоснованием применимости ИНС в качестве САУ служат теоремы Колмогорова–Арнольда и Хехт–Нильсена. Любой аналоговый регулятор может быть с заданной точностью создан с помощью сетей из нейроподобных элементов.

В качестве объекта управления выбран процесс задания соотношения «руда/вода» в стержневых мельницах секций мокрой магнитной сепарации АО ССГПО. Выбор обусловлен сильной зависимостью указанного параметра от физико-химических свойств подаваемой на измельчение смеси руд различных месторождений северного Казахстана и отсутствием формализованных закономерностей его задания.

На данных АО ССГПО было произведено аппроксимирование информационного пространства различными методами и разными архитектурами нейросетей. Сеть Radial Basis Function Network показала наилучшую точность и достаточно высокое, по сравнению с другими видами нейросетей и алгоритмов аппроксимации, время перенастройки.

Авторами разработана структура САУ с блоком адаптивной памяти на основе сети RBFN и ее программная реализация в среде визуального программирования SIMULINK.

По результатам анализа модели сделаны заключения. Моделирование произведено с погрешностью порядка 5%, данная точность достигнута при реализации 600 прогонов модели. Модель показывает значимое снижение энергопотребления и прирост производительности процесса измельчения при использовании разработанной адаптивной САУ по сравнению с имеющимися на предприятии АСУ.

Работа выполнена в рамках гранта №601 МОН РК по приоритету «Информационные и телекоммуникационные технологии» по направлению «Интеллектуальные системы и технологии».

- [1] *Зарубин М. Ю., Зарубина В. Р.* Использование искусственных нейроподобных сетей для управления технологическими процессами железо-обоганительного комплекса // Искусственный интеллект, Донецк. — 2013 (принята в печать).

## Нейросетевое и алгебраическое моделирование в распределенных волоконно-оптических измерительных системах

*Закасовская Елена Владимировна*

*Кульчин Юрий Николаевич*

zakasovskaya@inbox.ru, kulchin@iacp.dvo.ru

Владивосток, Дальневосточный федеральный университет,

Институт автоматизации и процессов управления ДВО РАН

Решение задачи оценки параметров распределенных физических полей (ФП), подверженных априорно неизвестным внешним воздействиям, имеет исключительно большое значение. В связи с этим возникает необходимость создания измерительных средств, обладающих распределенной в пространстве чувствительностью. Широкие возможности в подобных задачах открывает использование измерительных систем на основе распределенных волоконно-оптических измерительных сетей (ВОИС). В таких измерительных системах оптические сигналы, поступающие по волоконно-оптическим измерительным линиям (ИЛ), изменяются пропорционально величине внешнего физического воздействия. Сигнал на выходе ИЛ представляет собой линейный интеграл от функции распределения исследуемого параметра физического поля. Таким образом, возникает томографическая задача восстановления функций. Для типичных схем расположения ИЛ в ВОИС характерны сильная дискретизация по углу и по числу проекций, наличие нестандартных схем укладки ИЛ, большой объем данных, низкая скорость реконструкции, что не позволяет создавать быстродействующие измерительные системы.

В работе представлен новый эффективный метод реконструкции распределений параметров ФП в условиях неравномерной дискретизации томографических проекций, формируемых ВОИС; предложен метод устранения артефактов, основанный на алгебраическом и нейросетевом синтезе синопграмм; разработаны модели обработки неполных данных на основе комплексов нейронных сетей.

- [1] *Кульчин Ю. Н., Закасовская Е. В.* Использование последовательности комплексов нейронных сетей в распределенных волоконно-оптических измерительных системах // Проблемы управления. — 2011. — № 4. — С. 5–12. [http://pu.mtas.ru/archive/contents\\\_4\\\_2011.php](http://pu.mtas.ru/archive/contents\_4\_2011.php).

## Агрегирование и аппроксимация задержек для пар процессов в коммуникационной среде многопроцессорных систем

*Майсурадзе Арчил Ивериевич*<sup>1</sup>

*Сальников Алексей Николаевич*<sup>2</sup>

*Андреев Дмитрий Юрьевич*<sup>3</sup>

<sup>1</sup>maysuradze@cs.msu.su, <sup>2</sup>salnikov@cs.msu.su, <sup>3</sup>andreevd@cs.msu.su

Москва, Московский государственный университет имени М. В. Ломоносова

Для изучения коммуникационной среды многопроцессорных систем применяются различные подходы к сбору информации о передаче сообщений и её анализу. В докладе используется подход, основанный на измерении задержек при передаче с помощью специальных тестов.

Тесты создают огромный объем данных, пропорциональный количеству пар процессов, вариантов сообщений, режимов работы коммуникационной среды. Результаты анализа используются при создании и оптимизации параллельных программ, при функционировании планировщиков, для обнаружения аномалий коммуникационной среды и вычислительных узлов. В ряде случаев результаты анализа должны использоваться в режиме реального времени. Следовательно, требуется агрегирование собранной информации, которое позволит достичь компромисса между ресурсоёмкостью системы, скоростью и качеством работы.

В докладе рассматривается ситуация, когда анализируется информация о свойствах пар процессов. Предлагается провести кластеризацию пар процессов, что для современных суперкомпьютеров соответствует обработке десятков тысяч объектов и сотен миллионов расстояний. Предлагается мера сходства, которая группирует такие пары процессов, для которых задержка с ростом длины сообщения растёт похожим образом. Предполагается, что в корректно функционирующей коммуникационной среде число типов поведения задержки должно быть небольшим.

Сравниваются различные методы кластеризации и эвристики, ускоряющие эти методы. Эмпирически показано, что качество аппроксимации исходных данных остаётся приемлемым при использовании сверхбыстрых методов, которые не запрашивают большинство расстояний. Кроме того, эти методы допускают эффективную параллельную реализацию.

Работа поддержана грантом РФФИ № 13-01-00751.

- [1] Сальников А. Н., Майсурадзе А. И., Андреев Д. Ю., Костин Г. А. Кластеризация результатов тестирования коммуникационной среды многопроцессорных систем: единицы анализа, исследование методов, визуализация результатов // Вестник УГАТУ, Т. 16, № 6(51), Уфа: УГАТУ, 2012. — С. 149–157. [http://www.ugatu.ac.ru/publish/vu/stat/ugatu-2012-6\(51\)/22.pdf](http://www.ugatu.ac.ru/publish/vu/stat/ugatu-2012-6(51)/22.pdf).

## Применение процедуры парного выравнивания для разметки стыков на ультразвуковой дефектограмме рельсового пути

*Маленичев Антон Александрович*<sup>1</sup>

*Сулимова Валентина Вячеславовна*<sup>2</sup>

*Красоткина Ольга Вячеславовна*<sup>3</sup>

*Моттль Вадим Вячеславович*<sup>4</sup>

*Марков Анатолий Аркадиевич*<sup>5</sup>

<sup>1</sup>malenichev@mail.ru, <sup>2</sup>vsulimova@yandex.ru,

<sup>3</sup>o.v.krasotkina@yandex.ru, <sup>4</sup>vmottl@yandex.ru,

<sup>5</sup>info@radioavionica.ru

<sup>123</sup>Тула, <sup>4</sup>Москва, <sup>5</sup>Санкт-Петербург, <sup>123</sup>ФГБОУ ВПО ТулГУ, <sup>4</sup>ВЦ РАН,

<sup>5</sup>ОАО «Радиоавионика»

Дефектограмма представляет собой дискретный сигнал, зарегистрированный дефектоскопом при ультразвуковом исследовании рельсового пути. При наличии дефектов или каких-либо конструктивных отражателей (например, болтового соединения) на дефектограмме появляются линии разной формы, ориентации и продолжительности, в зависимости от вида объекта, оказавшегося на пути распространения сигнала.

Для снижения вероятности возникновения аварийных ситуаций на железной дороге важно не только верно идентифицировать дефекты рельсов, но и определять их местоположение на рельсовом пути.

В связи с этим актуальной является задача разметки дефектограммы, под которой в данной работе понимается выделение областей, соответствующих основным конструктивным элементам — рельсовым стыкам, с расстановкой характерных точек, указывающих непосредственно место соединения пары рельсов, а также начало и окончание каждой «зоны стыка».

Поскольку скорость движения дефектоскопа не постоянна, сигналы, полученные при исследовании даже одного и того же участка рельса в общем случае имеют различную длину и форму. Ввиду этого задача разметки решается на основе сравнения фрагментов исследуемой дефектограммы с эталонным фрагментом, представляющим типовой стык, с использованием специально разработанной меры несходства сигналов, основанной на парном выравнивании, и учитывающей особенности дефектограмм.

Работа поддержана грантами РФФИ № 12-07-13141, 12-07-13142

- [1] Маленичев А. А., Сулимова В. В., Красоткина О. В., Моттль В. В., Марков А. А. Применение процедуры парного выравнивания для разметки стыков на ультразвуковой дефектограмме рельсового пути // Известия ТулГУ. Технические науки. Вып. 9. Тула: Изд-во ТулГУ, 2013.

## Диагностика технического состояния подшипников на основе обнаружения локальных неоднородностей в вибросигнале

*Чувиллина Елена Владимировна*

*e.v.chuvilina@gmail.com*

Рыбинск, Рыбинский государственный авиационный технический университет имени П. А. Соловьева

В работе предложен алгоритм решения задачи диагностики технического состояния подшипников ГТД по вибросигналу на основе обнаружения в нем локальных неоднородностей. Используется постановка задачи обнаружения изменения свойств случайных процессов, поставленная В. В. Моттлем. Поскольку обнаружение изменений любой функции распределения или какой-либо иной вероятностной характеристики может быть сведено к обнаружению изменения математического ожидания в некоторой новой последовательности, сформированной из исходной, то задача обнаружения локальных неоднородностей в вибросигнале сводится к задаче обнаружения изменения свойств в диагностической последовательности.

В качестве диагностических признаков рассматриваются изменение фрактальной размерности, расстояние между матрицами зависимости величины приращения сигнала от величины сигнала, расстояние между векторами сноса на соседних блоках сигнала. Вводится критерий обнаружения локальной неоднородности на основе изменения фрактальной размерности. Для проведения эксперимента используется выборка вибросигналов подшипников ГТД, которая может быть разбита на три класса: неисправные ( $B$ ), кондиционные ( $C$ ) и необоснованно снятые ( $N$ ).

Для оценки ценности предлагаемых диагностических признаков рассчитана их информативность как отношение межклассового расстояния к внутриклассовому. По результатам вычислительных экспериментов можно выделить количество «положительных», «отрицательных», общее количество неоднородностей, расстояние между векторами сноса, расстояние между матрицами зависимости приращения от величины сигнала. Их информативность заметно выше, чем для СКО, полученного по сигналу ИВУ-1М, и спектральных признаков, определяемых с помощью комплекса МПС-200. Использование предложенных характеристик обеспечивает полную разделимость подшипников в предлагаемом признаковом пространстве, то есть их классификацию.

- [1] Чувиллина Е. В. Диагностика состояния подшипников на основе обнаружения локальных неоднородностей в вибросигнале // Машинное обучение и анализ данных. — 2013. <http://jmla.org/papers/doc/.pdf>.

## Прогноз вибрации высотных сооружений

Дергузов Аркадий Владимирович<sup>1</sup>

Махортых Сергей Александрович

<sup>1</sup>arkadid@list.ru

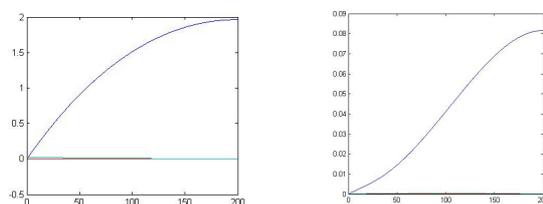
Пушино, ИМПБ РАН

Целями представляемой работы являлась разработка методов диагностики технического состояния строений на поверхности земли, подвергаемых техногенным, метеорологическим и прочим воздействиям. Предложены модели рассматриваемых систем, в том числе включающие средства по снижению внешних воздействий на сооружения. В основу методики диагностики состояния конструкций положены спектрально-аналитические подходы. В качестве признаков используются спектральные коэффициенты разложения сигналов в ортогональные ряды, а также импедансные параметры систем.

Метод применялся для оценки и прогнозирования передачи вибрации, вызванной техногенными, сейсмическими или метеорологическими факторами по структуре высотного здания на импульсное или периодическое акустическое воздействие (для схемы высотного здания и источник вибрации (линия метрополитена)). В качестве математической модели рассматриваемой системы (здания) была использована модель защемленной балки. Для уменьшения значений поля вибрации для случая наличия ветровой нагрузки по поверхности здания предлагается использовать резонатор, расположенный в верхней части здания и представляющий собой демпфированный груз на пружине.

Поле вибрации:  $EJ\ddot{u} + m\dot{u} = K\dot{u} + p(z, t) + F$ , где  $J$  — момент инерции поперечного сечения конструкции здания,  $m$  — погонная масса здания,  $K$  — сила упругости пружины,  $p(z, t)$  — сила ветра по длине здания,  $F$  — сила воздействия резонатора. На рисунках показано поле вибрации вдоль высоты здания для ветра, движущегося со скоростью 30 м/с с резонатором в верхней части балки и без резонатора для частоты 2 Гц.

При поддержке грантов РФФИ 11-07-00519, 11-07-00716, 13-01-00340.



- [1] Дергузов А. В., Махортых С. А. Резонансно-импедансный метод диагностики одномерных акустических систем // Контроль. Диагностика. №4. 2006.

## Программный комплекс для высокоточной аппроксимации многомерных временных рядов на базе вычислительного кластера ИМПБ РАН

*Рыкунов Станислав Дмитриевич*

*Панкратов Антон Николаевич*

*Устинин Михаил Николаевич*

stanislavrykunov@gmail.com

Пушино, Институт математических проблем биологии РАН

Многопроцессорный Linux-кластер ИМПБ РАН является ядром информационно-вычислительного комплекса Пушкинского научного центра [1]. После очередной модернизации в 2010–2013 гг., проведенной на средства РФФИ, производительность кластера достигла 1.5 ТФЛОПС. В настоящее время кластер состоит из вычислительных узлов двух типов: 9 двухпроцессорных узлов на базе процессоров Intel Xeon X5650 (6 ядер, 2.66ГГц); и 12 двухпроцессорных узлов на базе процессоров Intel Xeon E5620 (4 ядра, 2.40ГГц); на каждый узел приходится по 24 Гб оперативной памяти. Суммарное количество вычислительных ядер — 204, объем оперативной памяти — 504 Гб.

Программный комплекс для решения задач обработки и интеллектуального анализа сложных экспериментальных данных был создан и отлажен на примере многоканальных данных энцефалографии. Комплекс позволяет вычислять высокоточные разложения в ряды Фурье для последующего многомерного частотного анализа. Данные были представлены в виде массивов чисел двойной точности, размеры массивов определялись временем регистрации данных и числом измерительных каналов и составляли  $150000 \times 148$  или  $300000 \times 148$ . Для интерпретации программы использовалась среда GNU Octave. Для ускорения обработки данных были применены распараллеливание по реализациям. Был выбран узел кластера с 8 ядрами и для каждого набора исходных данных запущен свой вычислительный процесс. Среднее время исполнения программы для массива  $150000 \times 148$  составило 38 часов 40 минут, а для массива  $300000 \times 148$  — 156 часов 15 минут. Проверка точности аппроксимации показала, что отклонение между исходными и восстановленными данными не превышает 0.01%.

Работа поддержана грантами РФФИ № 13-07-00162 и № 11-07-00716.

- [1] *Лакно В. Д. и др.* Развитие информационно-коммуникационных технологий в Пушкинском научном центре РАН // Математическая биология и биоинформатика, 2012. — Т.7 № 2. С.529–544 [http://www.matbio.org/2012/Lakhno\\_7\\_529.pdf](http://www.matbio.org/2012/Lakhno_7_529.pdf).

## Облачные технологии и их применение в задачах вычислительной биологии

*Оплачко Екатерина Сергеевна*<sup>1</sup>

*Устинин Дмитрий Михайлович*<sup>2</sup>

*Устинин Михаил Николаевич*<sup>3</sup>

<sup>1</sup>oplachkoe@gmail.com, <sup>3</sup>Ustinin@impb.psn.ru

Пуццино, Институт математических проблем биологии РАН

Облачные вычисления представляют собой сервис, основанный на понятии подписки, где пользователь может получить доступ к аппаратным мощностям или конечному программному обеспечению удаленно. На текущий момент многие крупные компании (производители как оборудования, так и программного обеспечения) предоставляют свои облачные сервисы. Наиболее известные из них: Windows Azure, IBM Smart Business, Amazon Web Service (AWS). Облачные технологии обладают рядом преимуществ — это доступ к необходимой системе из любой точки мира, где есть Интернет, экономия на покупке и обслуживании аппаратных ресурсов и программного обеспечения. Для научного сообщества это возможность использовать аппаратные мощности крупных вычислительных центров. Вычислительная биология, с ее объемами обрабатываемых данных и масштабом вычислений, не является исключением. На текущий момент уже существует несколько проектов по предоставлению облачной инфраструктуры для исследовательских целей (Cloud BioLinux, Science Cloud). При этом не существует проектов, которые готовы предоставить биологам готовое программное обеспечение по моделированию работы биологической клетки в режиме Software as a Service (SaaS).

В Институте математических проблем биологии РАН разработан проект «Математическая клетка» ([www.mathcell.ru](http://www.mathcell.ru)), который представляет собой набор ресурсов и моделей для исследования работы биологической клетки. Данная система частично соответствует понятию SaaS в рамках Публичной подписки (Public Cloud). Однако требуются доработки в области мониторинга ресурсов, автоматизации масштабируемости и балансировки нагрузки.

ИМПБ РАН планирует развертывание полноценной облачной системы на своих аппаратных ресурсах, с использованием технологий виртуализации, при этом будет учтена необходимость балансировки нагрузки и легкой масштабируемости. Соответственно, все ресурсы проекта «Математическая клетка» будут перенесены в это облако.

- [1] *Оплачко Е. С., Устинин Д. М., Устинин М. Н.* Облачные технологии и их применение в задачах вычислительной биологии // Математическая биология и биоинформатика. Т. 8, выпуск 2, 2013 (статья принята к изданию).

## Содержание

<b>Композиционные методы анализа данных</b> . . . . .	5
<i>Дьяконов А. Г.</i>	
Решение задач анализа данных, основанное на линейной комбинации деформаций . . . . .	5
<i>Хачай М. Ю., Поберий М. И.</i>	
Схема бустинга в задачах коллективного обучения . . . . .	6
<i>Дюкова Е. В., Любимцева М. М., Прокофьев П. А.</i>	
Логические корректоры в задачах классификации по прецедентам . . . . .	7
<i>Шибзухов З. М.</i>	
О потоечно корректных операциях над алгоритмами распознавания и прогнозирования . . . . .	8
<i>Матросов В. Л., Шибзухов З. М.</i>	
Об агрегированно корректных операциях над алгоритмами . . . . .	9
<i>Бериков В. Б.</i>	
Выбор оптимальных весов в коллективе алгоритмов кластерного анализа . . . . .	10
<i>Суворов М. А., Майсурадзе А. И.</i>	
Методы агрегирования метрических описаний . . . . .	11
<i>Романенко А. А.</i>	
Агрегирование адаптивных алгоритмов прогнозирования при линейной несимметричной функции потерь . . . . .	12
<b>Линейные и метрические методы восстановления зависимостей</b> . . . . .	13
<i>Моттль В. В., Середин О. С.</i>	
Линейные методы восстановления зависимостей по эмпирическим данным в метрических пространствах . . . . .	13
<i>Двоенко С. Д., Пшеничный Д. О.</i>	
Метрическая коррекция матриц парных сравнений . . . . .	14
<i>Двоенко С. Д., Пшеничный Д. О.</i>	
О локализации отрицательных собственных значений . . . . .	15
<i>Губарева А. А., Ларин А. О., Середин О. С., Сулимова В. В.</i>	
Поиск наибольшей гиперполости в линейном пространстве данных . . . . .	16
<i>Абрамов В. И., Середин О. С., Моттль В. В.</i>	
Логистическая регрессия в метрических пространствах . . . . .	17

<i>Турков П. А., Красоткина О. В., Моттль В. В.</i> Байесовская логистическая регрессия в задаче обучения распознаванию образов при смещении решающего правила . . . . .	18
<i>Разин Н. А., Черноусова Е. О., Красоткина О. В., Моттль В. В.</i> Применение Машины Релевантных Объектов в задачах восстановления числовых зависимостей . . . . .	19
<i>Разин Н. А., Моттль В. В.</i> Численная реализация алгоритмов селективного комбинирования разнородных представлений объектов в задачах распознавания образов . . . . .	20
<i>Красоткина О. В., Моттль В. В., Разин Н. А., Черноусова Е. О.</i> Беспереборный метод скользящего контроля для верификации квадратично-модульной регрессионной модели . . . . .	21
<b>Дискретно-логические методы классификации . . . . .</b>	<b>22</b>
<i>Неделько В. М.</i> Исследование погрешности оценок скользящего экзамена . . . . .	22
<i>Ишкина Ш. Х., Ивахненко А. А.</i> Комбинаторные оценки переобучения пороговых решающих правил . . . . .	23
<i>Генрихов И. Е.</i> Снижение переобученности распознающих алгоритмов на основе полных решающих деревьев . . . . .	24
<i>Прокашева О. В.</i> Повышение эффективности алгоритма классификации на основе анализа формальных понятий . . . . .	25
<i>Цыганкова И. А.</i> Эволюционный метод классификация биологических объектов . . . . .	26
<i>Бериков В. Б., Пестунов И. А., Герасимов М. К.</i> Метод анализа ансамбля многомерных разнотипных временных рядов и его применение при исследовании заболеваемости клещевым энцефалитом . . . . .	27
<i>Янковская А. Е., Китлер С. В.</i> Интеллектуальный анализ данных и знаний по стентированию коронарных артерий . . . . .	28
<i>Сенько О. В., Кузнецова А. В., Костомарова И. В.</i> Использование модифицированной версии метода оптимальных достоверных разбиений для анализа клинических данных . . . . .	29

<i>Теклина Л. Г., Котельников И. В., Гельфер И. С.</i> Синтез простейших нелинейных систем квазиинвариантного управления с заданными свойствами методами распознавания образов . . . . .	30
<b>Сложность вычислений и оптимизация . . . . .</b>	<b>31</b>
<i>Кельманов А. В.</i> О некоторых моделях «обучения» компьютера, индуцирующих NP-трудные задачи дискретной оптимизации . . . . .	31
<i>Ерёмин И. И., Гимади Э. Х., Кельманов А. В., Пяткин А. В., Хачай М. Ю.</i> 2-приближенный алгоритм для одной задачи поиска «похожих» объектов . . . . .	32
<i>Кельманов А. В., Романченко С. М.</i> FPTAS для одной труднорешаемой задачи поиска подмножества векторов . . . . .	33
<i>Галашов А. Е., Кельманов А. В.</i> 2-приближенный алгоритм для одной задачи поиска семейства непересекающихся подмножеств векторов . . . . .	34
<i>Кельманов А. В., Хандеев В. И.</i> Рандомизированный алгоритм для одной NP-трудной задачи кластерного анализа . . . . .	35
<i>Кельманов А. В., Пяткин А. В.</i> О сложности одной задачи кластерного анализа . . . . .	36
<i>Кельманов А. В., Хамидуллин С. А.</i> 2-приближенный полиномиальный алгоритм для одной NP-трудной задачи кластеризации последовательности . . . . .	37
<i>Кельманов А. В., Михайлова Л. В.</i> Задача суммирования элементов двух последовательностей . . . . .	38
<i>Зухба А. В.</i> Оценка вычислительной сложности задачи монотонизации выборки . . . . .	39
<i>Боровых Н. И., Красоткина О. В.</i> Применение алгоритма поиска с запретами в задаче автоматизированного составления оптимального штатного расписания . . . . .	40
<i>Гасников А. В., Бондаренко А. С.</i> Об одной вариации метода экспоненциального взвешивания для решения задач стохастической онлайн оптимизации . . . . .	41

<i>Заботин И. Я., Яруллин Р. С.</i> Об одном методе отсечений с отбрасыванием отсекающих плоскостей . . . . .	42
<b>Обработка изображений . . . . .</b>	<b>43</b>
<i>Харинов М. В.</i> Иерархические и оптимальные приближения цифрового изображения . . . . .	43
<i>Орлов А. А., Астафьев А. В.</i> Алгоритм цифровой локализации изображений символьных меток на основе одномерного дифференциально-интегрального исчисления яркости . . . . .	44
<i>Роженцов А. А., Морозовский К. В.</i> Адаптивное модифицированное обобщённое трехмерное преобразование Хафа . . . . .	45
<i>Мнухин В. Б.</i> Цифровые изображения на комплексном дискретном торе . . . . .	46
<i>Федотов Н. Г., Голдуева Д. А.</i> Анализ трехмерных текстур с позиции стохастической геометрии и функционального анализа . . . . .	47
<i>Жизняков А. Л., Привезенцев Д. Г.</i> Использование характера распределения самоподобия в качестве признака цифрового изображения в задаче классификации . . . . .	48
<i>Федосеев В. А.</i> Оптимизация ядра диффузии ошибки для растривания полутоновых изображений . . . . .	49
<i>Мурашов Д. М., Березин А. В., Иванова Е. Ю.</i> Формирование признакового описания фактуры картин . . . . .	50
<b>Анализ изображений . . . . .</b>	<b>51</b>
<i>Шаповалов Р. В., Ветров Д. П., Коли П.</i> Учёт далекодействующих зависимостей в задаче семантической сегментации трёхмерных облаков точек с помощью последовательной классификации . . . . .	51
<i>Осокин А. А., Коли П., Йегелька Ш.</i> Иерархическое марковское случайное поле для задачи сегментации изображений . . . . .	52

<i>Янгель Б. К., Ветров Д. П.</i> Обучение модели формы для сегментации изображений по слабо размеченным данным . . . . .	53
<i>Грачева И. А., Копылов А. В.</i> Адаптивный параметрический алгоритм обработки изображений	54
<i>Пестунов И. А., Мельников П. В., Сняевский Ю. Н.</i> Ансамблевые алгоритмы сегментации спутниковых изображений по спектральным и пространственным признакам . . . . .	55
<b>Анализ видеопоследовательностей . . . . .</b>	<b>56</b>
<i>Чочиа П. А.</i> Анализ видеопоследовательностей, формируемых капилляроскопом . . . . .	56
<i>Хашин С. И.</i> Динамическая сегментация пары кадров . . . . .	57
<i>Малин И. К., Вишняков Б. В.</i> Метод автоматической сегментации человеческих фигур на изображениях . . . . .	58
<i>Вишняков Б. В., Малин И. К., Егоров А. И.</i> Модель статического фона с использованием дескрипторов LBP и LTP . . . . .	59
<b>Распознавание изображений . . . . .</b>	<b>60</b>
<i>Гарбук С. В.</i> Конкурс Фонда перспективных исследований на лучший прототип технологии распознавания образов . . . . .	60
<i>Горбачев В. С.</i> Отбор и обучение бинарных классификаторов в задаче распознавания лиц . . . . .	61
<i>Чинаев Н. Н., Матвеев И. А.</i> Определение точной границы зрачка . . . . .	62
<i>Нагапетян В. Э.</i> Распознавание жестов ручных азбук глухонемых . . . . .	63
<i>Лебедев Л. И.</i> Метод целевого нахождения согласованных описаний при распознавании объектов изображений . . . . .	64
<i>Каржищенко А. Н., Мнухин В. Б.</i> Восстановление симметричности точек на изображениях объектов с отражательной симметрией . . . . .	65

<i>Кривецкий А. В.</i>	
Обнаружение трехмерных групп точечных объектов на основе анализа графа иерархической группировки объемной сцены . . .	66
<i>Ларин А. О., Середин О. С., Кондрашов В. В.</i>	
Комбинирование одноклассовых классификаторов в задачах сегментации изображений, содержащих объекты топологии печатной платы . . . . .	67
<b>Анализ формы изображений . . . . .</b>	<b>68</b>
<i>Пытьев Ю. П.</i>	
Косые проекторы и относительные формы в морфологии изображений . . . . .	68
<i>Пытьев Ю. П., Нагорный Ю. М.</i>	
Морфологический метод выделения неизвестного объекта на изображении, полученном при неизвестном освещении . . . . .	69
<i>Визильтер Ю. В., Рубис А. Ю.</i>	
Морфологическое сравнение образов по сложности . . . . .	70
<i>Визильтер Ю. В., Сидякин С. В.</i>	
Параметрические и морфологические спектры . . . . .	71
<i>Жукова К. В., Рейер И. А.</i>	
Параметрический дескриптор формы и связность базового скелета	72
<i>Кушнир О. А., Середин О. С.</i>	
Построение функции ширины скелета бинарного изображения на основе параметрического описания многочленами Лежандра	73
<i>Кушнир О. А., Середин О. С., Степанов А. В.</i>	
Экспериментальное исследование параметров регуляризации и аппроксимации скелетных графов бинарных изображений . . .	74
<i>Ланге М. М., Ганебных С. Н.</i>	
Иерархические структуры данных и решающие алгоритмы для классификации изображений . . . . .	75
<i>Наумов А. С., Роженицов А. А.</i>	
Синтез и анализ алгоритма распознавания контурных сигналов с применением комплекснозначных нейронных сетей . . . . .	76
<i>Захаров А. А., Жизняков А. Л.</i>	
Автоматическая реконструкция трехмерных объектов по ортогональным проекциям . . . . .	77
<i>Анциперов В. Е., Евсеев О. В.</i>	
Метод параметрической реконструкции 3-D плотности распределения облака точек, заданного серией параллельных 2-D сечений	78

<b>Приложения: биология и медицина</b> . . . . .	79
<i>Обухов Ю. В., Королев М. С., Обухов К. Ю.</i>	
О количественной оценке нестационарности ритмов электроэнцефалограмм в норме и при паркинсонизме в ранних стадиях . . . . .	79
<i>Королев М. С., Обухов Ю. В., Обухов К. Ю., Сушкова О. С.</i>	
Совместный анализ частотно-временных характеристик сигналов электроэнцефалографии, электромиографии и механического тремора при болезни Паркинсона в ранних стадиях . . . . .	80
<i>Дергузов А. В., Махортыл С. А., Лыжко Е. В.</i>	
Классификация и распознавание образов в данных магнитной энцефалографии . . . . .	81
<i>Манило Л. А., Немирко А. П., Саламонова И. С.</i>	
Автоматический анализ формы спирографических петель по их сигнатурам . . . . .	82
<b>Приложения: химия и биоинформатика</b> . . . . .	83
<i>Торшин И. Ю., Рудаков К. В.</i>	
О комбинаторном тестировании условия локальной полноты связанных размеченных графов в применении к задачам хемоинформатики . . . . .	83
<i>Прохоров Е. И., Кумсков М. И.</i>	
Двухфазная схема распознавания в задаче «структура–свойство»	84
<i>Панкратов А. Н., Пятков М. И., Руднев В. Р., Куликова Л. И.</i>	
Спектральный подход к распознаванию повторов в геномах и белках . . . . .	85
<i>Панкратов А. Н., Руднев В. Р., Куликова Л. И., Дедус Ф. Ф., Тихонов Д. А., Ефимов А. В.</i>	
Распознавание и анализ устойчивости структурных мотивов типа $\alpha$ - $\alpha$ -уголок в глобулярных белках . . . . .	86
<b>Приложения: анализ текстов и веба</b> . . . . .	87
<i>Ботов П. В., Вьючнов Д. В., Корольков М. Е., Песков Н. В., Суровенко Н. С., Христанков А. С., Царьков С. В., Чехович Ю. В.</i>	
Анализ структуры зависимостей в корпусе взаимосвязанных текстов с применением системы «Антиплагиат» и интеллектуального анализа данных на примере диссертаций в области исторических наук . . . . .	87
<i>Воронцов К. В.</i>	
Аддитивная регуляризация тематических моделей . . . . .	88

<i>Потапенко А. А.</i>	
Разреживание вероятностных тематических моделей . . . . .	89
<i>Царёв Д. В., Машечкин И. В., Петровский М. И.</i>	
Методы вычисления релевантности фрагментов текста на основе тематических моделей в задаче автоматического аннотирования . . . . .	90
<i>Лексин В. А., Николенко С. И.</i>	
Извлечение тегов веб-страниц с частичным обучением . . . . .	91
<i>Нгуен Т. Т., Поленова Е. А., Красоткина О. В., Моттль В. В.</i>	
Байесовский подход к оцениванию факторов, влияющих на положение сайта в результатах поискового запроса . . . . .	92
<i>Кудинов М. С.</i>	
Частичный синтаксический разбор с помощью условных случайных полей . . . . .	93
<i>Чувиллин К. В.</i>	
Использование правил со сложной структурой для коррекции документов в формате LaTeX . . . . .	94
<b>Прикладные системы . . . . .</b>	<b>95</b>
<i>Мандрикова О. В., Соловьев И. С.</i>	
Вейвлет-технология обработки и анализа геомагнитных данных . . . . .	95
<i>Мандрикова О. В., Глушкова Н. В., Полозов Ю. А.</i>	
Моделирование и анализ параметров ионосферы на основе совмещения вейвлет-преобразования и авторегрессионных моделей . . . . .	96
<i>Кузнецов А. В., Мясников В. В.</i>	
Построение вычислительной процедуры комплексной проверки подлинности данных ДЗЗ . . . . .	97
<i>Зарубин М. Ю., Зарубина В. Р.</i>	
Использование искусственных нейрноподобных сетей для управления технологическими процессами железо-обогатительного комплекса . . . . .	98
<i>Закасовская Е. В., Кульчин Ю. Н.</i>	
Нейросетевое и алгебраическое моделирование в распределенных волоконно-оптических измерительных системах . . . . .	99
<i>Майсурадзе А. И., Сальников А. Н., Андреев Д. Ю.</i>	
Агрегирование и аппроксимация задержек для пар процессов в коммуникационной среде многопроцессорных систем . . . . .	100

---

<i>Маленичев А. А., Сулимова В. В., Красоткина О. В., Моттль В. В., Марков А. А.</i>	
Применение процедуры парного выравнивания для разметки стыков на ультразвуковой дефектограмме рельсового пути . . .	101
<i>Чувилина Е. В.</i>	
Диагностика технического состояния подшипников на основе обнаружения локальных неоднородностей в вибросигнале . . . . .	102
<i>Дергузов А. В., Махортых С. А.</i>	
Прогноз вибрации высотных сооружений . . . . .	103
<i>Рыкунов С. Д., Панкратов А. Н., Устинин М. Н.</i>	
Программный комплекс для высокоточной аппроксимации многомерных временных рядов на базе вычислительного кластера ИМПБ РАН . . . . .	104
<i>Оплачко Е. С., Устинин Д. М., Устинин М. Н.</i>	
Облачные технологии и их применение в задачах вычислительной биологии . . . . .	105

**Авторский указатель**

- А**  
Абрамов В. И. .... 17  
Андреев Д. Ю. .... 100  
Анциперов В. Е. .... 78  
Астафьев А. В. .... 44
- Б**  
Березин А. В. .... 50  
Бериков В. Б. .... 10, 27  
Бондаренко А. С. .... 41  
Боровых Н. И. .... 40  
Ботов П. В. .... 87
- В**  
Ветров Д. П. .... 51, 53  
Визильтер Ю. В. .... 70, 71  
Вишняков Б. В. .... 58, 59  
Воронцов К. В. .... 88  
Вьючнов Д. В. .... 87
- Г**  
Галашов А. Е. .... 34  
Ганебных С. Н. .... 75  
Гарбук С. В. .... 60  
Гасников А. В. .... 41  
Гельфер И. С. .... 30  
Генрихов И. Е. .... 24  
Герасимов М. К. .... 27  
Гимади Э. Х. .... 32  
Глушкова Н. В. .... 96  
Голдуева Д. А. .... 47  
Горбацевич В. С. .... 61  
Грачева И. А. .... 54  
Губарева А. А. .... 16
- Д**  
Двоенко С. Д. .... 14, 15  
Дедус Ф. Ф. .... 86  
Дергузов А. В. .... 81, 103  
Дьяконов А. Г. .... 5
- Дюкова Е. В. .... 7
- Е**  
Евсеев О. В. .... 78  
Егоров А. И. .... 59  
Ерёмин И. И. .... 32  
Ефимов А. В. .... 86
- Ж**  
Жизняков А. Л. .... 48, 77  
Жукова К. В. .... 72
- З**  
Заботин И. Я. .... 42  
Закасовская Е. В. .... 99  
Зарубин М. Ю. .... 98  
Зарубина В. Р. .... 98  
Захаров А. А. .... 77  
Зухба А. В. .... 39
- И**  
Иванова Е. Ю. .... 50  
Ивахненко А. А. .... 23  
Ишкина Ш. Х. .... 23
- Й**  
Йегелька Ш. .... 52
- К**  
Каркищенко А. Н. .... 65  
Кельманов А. В. .... 31–38  
Китлер С. В. .... 28  
Коли П. .... 51, 52  
Кондрашов В. В. .... 67  
Копылов А. В. .... 54  
Королев М. С. .... 79, 80  
Корольков М. Е. .... 87  
Костомарова И. В. .... 29  
Котельников И. В. .... 30

Красоткина О. В. 18, 19, 21, 40,  
92, 101  
Кревецкий А. В. .... 66  
Кудинов М. С. .... 93  
Кузнецов А. В. .... 97  
Кузнецова А. В. .... 29  
Куликова Л. И. .... 85, 86  
Кульчин Ю. Н. .... 99  
Кумсков М. И. .... 84  
Кушнир О. А. .... 73, 74

**Л**

Ланге М. М. .... 75  
Ларин А. О. .... 16, 67  
Лебедев Л. И. .... 64  
Лексин В. А. .... 91  
Лыжко Е. В. .... 81  
Любимцева М. М. .... 7

**М**

Майсурадзе А. И. .... 11, 100  
Маленичев А. А. .... 101  
Малин И. К. .... 58, 59  
Мандрикова О. В. .... 95, 96  
Манило Л. А. .... 82  
Марков А. А. .... 101  
Матвеев И. А. .... 62  
Матросов В. Л. .... 9  
Махортых С. А. .... 81, 103  
Машечкин И. В. .... 90  
Мельников П. В. .... 55  
Михайлова Л. В. .... 38  
Мнухин В. Б. .... 46, 65  
Морозовский К. В. .... 45  
Моттль В. В. . 13, 17–21, 92, 101  
Мурашов Д. М. .... 50  
Мясников В. В. .... 97

**Н**

Нагапетян В. Э. .... 63  
Нагорный Ю. М. .... 69  
Наумов А. С. .... 76

Нгуен Т. Т. .... 92  
Неделько В. М. .... 22  
Немирко А. П. .... 82  
Николенко С. И. .... 91

**О**

Обухов К. Ю. .... 79, 80  
Обухов Ю. В. .... 79, 80  
Оплачко Е. С. .... 105  
Орлов А. А. .... 44  
Осокин А. А. .... 52

**П**

Панкратов А. Н. .... 85, 86, 104  
Песков Н. В. .... 87  
Пестунов И. А. .... 27, 55  
Петровский М. И. .... 90  
Поберий М. И. .... 6  
Поленова Е. А. .... 92  
Полозов Ю. А. .... 96  
Потапенко А. А. .... 89  
Привезенцев Д. Г. .... 48  
Прокашева О. В. .... 25  
Прокофьев П. А. .... 7  
Прохоров Е. И. .... 84  
Пшеничный Д. О. .... 14, 15  
Пытьев Ю. П. .... 68, 69  
Пяткин А. В. .... 32, 36  
Пятков М. И. .... 85

**Р**

Разин Н. А. .... 19–21  
Рейер И. А. .... 72  
Роженцов А. А. .... 45, 76  
Романенко А. А. .... 12  
Романченко С. М. .... 33  
Рубис А. Ю. .... 70  
Рудаков К. В. .... 83  
Руднев В. Р. .... 85, 86  
Рыкунов С. Д. .... 104

**С**

Саламонова И. С. .... 82

Сальников А. Н. .... 100  
Сенько О. В. .... 29  
Середин О. С. 13, 16, 17, 67, 73,  
74  
Сидякин С. В. .... 71  
Синявский Ю. Н. .... 55  
Соловьев И. С. .... 95  
Степанов А. В. .... 74  
Суворов М. А. .... 11  
Сулимова В. В. .... 16, 101  
Суровенко Н. С. .... 87  
Сушкова О. С. .... 80

**Т**

Теклина Л. Г. .... 30  
Тихонов Д. А. .... 86  
Торшин И. Ю. .... 83  
Турков П. А. .... 18

**У**

Устинин Д. М. .... 105  
Устинин М. Н. .... 104, 105

**Ф**

Федосеев В. А. .... 49  
Федотов Н. Г. .... 47

**Х**

Хамидуллин С. А. .... 37

Хандеев В. И. .... 35  
Харинов М. В. .... 43  
Хачай М. Ю. .... 6, 32  
Хашин С. И. .... 57  
Хританков А. С. .... 87

**Ц**

Царёв Д. В. .... 90  
Царьков С. В. .... 87  
Цыганкова И. А. .... 26

**Ч**

Черноусова Е. О. .... 19, 21  
Чехович Ю. В. .... 87  
Чинаев Н. Н. .... 62  
Чочиа П. А. .... 56  
Чувиллин К. В. .... 94  
Чувилина Е. В. .... 102

**Ш**

Шаповалов Р. В. .... 51  
Шибзухов З. М. .... 8, 9

**Я**

Янгель Б. К. .... 53  
Янковская А. Е. .... 28  
Яруллин Р. С. .... 42

*Научное издание*

МАТЕМАТИЧЕСКИЕ МЕТОДЫ  
РАСПОЗНАВАНИЯ ОБРАЗОВ

Тезисы докладов  
16-й Всероссийской конференции

Напечатано с готового оригинал-макета

Сдано в набор 01.08.13. Подписано в печать 23.09.13.  
Печать цифровая. Бумага офсетная.  
Формат 70×100 1/16. Усл.-печ. л. 9,5. Уч.-изд. л. 10,5.  
Тираж 200 экз. Заказ

Издательство «ТОРУС ПРЕСС»  
121614, г. Москва, ул. Крылатская, 29-1-43  
e-mail: [torus@torus-press.ru](mailto:torus@torus-press.ru)  
<http://www.torus-press.ru>

Отпечатано в цифровой типографии «Буки Веди»  
на оборудовании Konica Minolta  
ООО «Ваш полиграфический партнер»  
г. Москва, ул. Ильменский пр-д, д. 1, корп. 6