

РОССИЙСКАЯ АКАДЕМИЯ НАУК
ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
«ИНФОРМАТИКА И УПРАВЛЕНИЕ» РАН
ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР ИМ. А. А. ДОРОДНИЦЫНА РАН
МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ
БАЛТИЙСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ ИМ. И. КАНТА

ПРИ ПОДДЕРЖКЕ

РОССИЙСКОГО ФОНДА ФУНДАМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ,
КОМПАНИЙ O1 GROUP, ФОРЕКСИС И ЦСПИР

Математические методы распознавания образов

ММРО-17

Тезисы докладов
17-й Всероссийской конференции
с международным участием

ТОРУС
ПРЕСС  МОСКВА
2015

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

ББК 22.1:32.973.26-018.2

М 34

Математические методы распознавания образов: Тезисы докладов 17-й Всероссийской конференции с международным участием, г. Светлогорск, 2015 г. — М.: ТОРУС ПРЕСС, 2015. — 274 с.
ISBN 978-5-94588-188-4

В сборнике представлены тезисы докладов 17-й Всероссийской конференции «Математические методы распознавания образов», проводимой Вычислительным центром им. А. А. Дородницына ФИЦ ИУ РАН, Московским физико-техническим институтом и Балтийским федеральным университетом им. И. Канта.

Конференция проводится регулярно, начиная с 1983 г., и является представительным научным форумом в области интеллектуального анализа данных, машинного обучения, распознавания образов, анализа изображений, обработки сигналов, дискретного анализа.

Организационный комитет ММРО-17 выражает особую благодарность РФФИ и компаниям O1 Group, Форексис и ЦСПиР, оказавшим неоценимую помощь при подготовке и проведении конференции.

ББК 22.1:32.973.26-018.2

ISBN 978-5-94588-188-4

© Авторы докладов, 2015

© ФИЦ ИУ РАН, 2015

UDK 004.85+004.89+004.93+519.2+519.25+519.7
BBK 22.1:32.973.26-018.2

Mathematical Methods for Pattern Recognition: Book of abstract of the 17th All-Russian Conference with International Participation, Svetlogorsk, 2015. — Moscow: TORUS PRESS, 2015. — 274 p.
ISBN 978-5-94588-188-4

The volume contains the abstracts of the 17th All-Russian Conference “Mathematical Methods for Pattern Recognition.” The conference is organized by Dorodnicyn Computing Centre FRC CSC RAS, Moscow Institute of Physics and Technology, and Immanuel Kant Baltic Federal University. The conference has been held biennially since 1983. It is one of the most recognizable scientific forums on data mining, machine learning, pattern recognition, image analysis, signal processing, and discrete analysis.

The Organizing Committee of MMPR-17 is grateful to the Russian Foundation for Basic Research, O1 Group, Forecsys Co. and CFRS Co. for providing assistance in the conference preparation and execution.

BBK 22.1:32.973.26-018.2

ISBN 978-5-94588-188-4

© Authors of the abstracts, 2015
© FRC CSC RAS, 2015

Оргкомитет

Председатель: Журавлев Юрий Иванович, *академик РАН*

Ученый секретарь: Чехович Юрий Викторович *к.ф.-м.н.*

Борисова Татьяна Игоревна

Громов Андрей Николаевич

Ивахненко Андрей Александрович, *к.ф.-м.н.*

Инякин Андрей Сергеевич, *к.ф.-м.н.*

Ишкина Шаура Хабировна

Помазкова Евгения Владимировна

Татарчук Александр Игоревич, *к.ф.-м.н.*

Программный комитет

Сопредседатели: Рудаков Константин Владимирович, *чл.-корр. РАН*

Зорин Денис Николаевич, *проф.*

Ученый секретарь: Стрижов Вадим Викторович, *д.ф.-м.н.*

Воронцов Константин Вячеславович, *д.ф.-м.н.*

Гимади Эдуард Хайрутдинович, *д.ф.-м.н.*

Горнов Александр Юрьевич, *д.т.н.*

Громова Ольга Алексеевна, *д.м.н.*

Двоенко Сергей Данилович, *д.ф.-м.н.*

Донской Владимир Иосифович, *д.ф.-м.н.*

Кельманов Александр Васильевич, *д.ф.-м.н.*

Местецкий Леонид Моисеевич, *д.т.н.*

Моттль Вадим Вячеславович, *д.т.н.*

Осипов Геннадий Семенович, *д.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.*

Чуличков Алексей Иванович, *д.ф.-м.н.*

Хачай Михаил Юрьевич, *д.ф.-м.н.*

Шананин Александр Алексеевич, *д.ф.-м.н.*

Рецензенты

Адуенко А. А.
Бахтеев О. Ю.
Борисова И. А.
Гасников А. В.
Гнеушев А. Н.
Гороховский К. Ю.
Дьяконов А. Г.
Животовский Н. К.
Игнатъев В. Ю.
Инякин А. С.
Ишкина Ш. Х.
Каркищенко А. Н.
Катруца А. М.
Красоткина О. В.
Крымова Е. А.

Кудинов М. С.
Кузнецов М. П.
Кузнецова М. В.
Ланге М. М.
Майсурадзе А. И.
Максимов Ю. В.
Матвеев И. А.
Местецкий Л. М.
Мотренко А. П.
Мурашов Д. М.
Неделько В. М.
Новик В. П.
Одиноких Г. А.
Панов А. И.
Панов М. Е.

Рейер И. А.
Сенько О. В.
Середин О. С.
Скипор К. С.
Стрижов В. В.
Сулимова В. В.
Талипов К. И.
Торшин И. Ю.
Трёкин А. Н.
Турдаков Д. Ю.
Фрей А. И.
Хачай М. Ю.
Черепанов Е. В.
Чуличков А. И.

Organizing Committee

Chair: Yury Zhuravlev, *acad. of RAS*

Secretary: Yury Chekhovich, *C.Sc.*

Tatyana Borisova
Andrey Gromov
Andrey Inyakin, *C.Sc.*
Shauro Ishkina
Andrey Ivakhnenko, *C.Sc.*
Evgeniya Pomazkova
Aleksandr Tatarchuk, *C.Sc.*

Program Committee

co-Chairs: Konstantin Rudakov, *corr. member of RAS*
Denis Zorin, *Prof.*

Secretary: Vadim Strijov, *D.Sc.*

Aleksey Chulichkov, *D.Sc.*
Vladimir Donskoy, *D.Sc.*
Sergey Dvoenko, *D.Sc.*
Edward Gimadi, *D.Sc.*
Alexander Gornov, *D.Sc.*
Olga Gromova, *D.Sc.*
Alexander Kel'manov, *D.Sc.*
Michael Khachay, *D.Sc.*
Leonid Mestetskiy, *D.Sc.*
Vadim Mottl, *D.Sc.*
Gennadiy Osipov, *D.Sc.*
Yury Pytiev, *D.Sc.*
Ivan Reyer, *C.Sc.*
Vladimir Ryazanov, *D.Sc.*
Aleksandr Shanenin, *D.Sc.*
Viktor Soyfer, *D.Sc.*
Mikhail Ustinin, *D.Sc.*
Konstantin Vorontsov, *D.Sc.*

Reviewers

Aduenko A.
Bakhteev O.
Borisova I.
Cherepanov E.
Chulichkov A.
D'yakonov A.
Frei O.
Gasnikov A.
Gneushev A.
Gorokhovskiy K.
Ignat'ev V.
Inyakin A.
Ishkina Sh.
Karkishchenko A.
Katrutsa A.

Khachay M.
Krasotkina O.
Krymova E.
Kudinov M.
Kuznetsov M.
Kuznetsova M.
Lange M.
Maksimov Yu.
Matveev I.
Maysuradze A.
Mestetskiy L.
Motrenko A.
Murashov D.
Nedelko V.
Novik V.

Odinokikh G.
Panov A.
Panov M.
Reyer I.
Senko O.
Seredin O.
Skipor K.
Strijov V.
Sulimova V.
Talipov K.
Torshin I.
Trekin A.
Turdakov D.
Zhivotovskiy N.

Краткое оглавление

Композиционные методы анализа данных	10
Линейные и метрические методы восстановления зависимостей	48
Дискретно-логические методы классификации	68
Сложность вычислений и оптимизация	84
Обработка изображений	120
Анализ изображений	142
Анализ видеопоследовательностей	152
Распознавание изображений	170
Анализ формы изображений	174
Анализ сигналов	190
Приложения: биология и медицина	202
Приложения: анализ текстов	220
Прикладные системы	232
Содержание	242
Авторский указатель	265

Brief contents

Compositional Data Mining Methods	10
Linear and Metric Learning Methods	48
Rule-Based Classification	68
Computational Complexity and Optimization	84
Image Processing	120
Image Analysis	142
Video Analysis	152
Image Recognition	170
Images Shape Analysis	174
Signal Analysis	190
Applications: Medicine and Biology	202
Applications: Text Mining	220
Application Systems	232
Contents	242
Author index	269

Анализ пространства параметров в задачах выбора мультимodelей

Адуенко Александр Александрович^{1,*} aduenko1@gmail.com
Стрижов Вадим Викторович^{1,2} strijov@ccas.ru

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Вычислительный центр им. А. А. Дородницына РАН

Рассматривается задача выбора мультимodelей при построении modelей в задачах двухклассовой классификации. Мультимodelи являются интерпретируемым обобщением случая одной modelи, позволяющим учитывать неоднородности в данных. Признаковые пространства modelей в мультимodelи могут не совпадать. Кроме того, мультимodelь может содержать большое число близких modelей, что ведет к низкому качеству прогноза и отсутствию интерпретируемости. Для решения этой проблемы предлагается метод статистического сравнения modelей для прореживания мультимodelи. Вводится понятие адекватной мультимodelи, т. е. мультимodelи, все modelи в которой являются попарно статистически различимыми. Для статистического сравнения modelей предлагается ввести функцию близости между апостериорными распределениями параметров modelей. Такая функция должна быть определена для случая пары распределений с несовпадающими носителями, а также не различать два распределения, одно из которых является малоинформативным. Показано, что дивергенция Кульбака–Лейблера, расстояния Дженсона–Шеннона, Хеллингера, Бхаттачарая не удовлетворяют этому требованию. Предлагается функция близости для пары распределений, которая удовлетворяет этим требованиям. Исследованы асимптотические свойства распределения введенной функции близости в условиях истинности гипотезы о совпадении modelей. С помощью статистических свойств распределения введенной функции близости получены оценки на максимальное количество попарно различимых modelей в мультимodelи для выборки фиксированного размера [1].

Работа поддержана грантом РФФИ № 14-07-31205.

- [1] *Адуенко А. А., Стрижов В. В.* Совместный выбор объектов и признаков в задачах многоклассовой классификации коллекции документов // *Инфокоммуникационные технологии*, 2014. № 1. С. 47–53.

Features space analysis for multimodel selection

Aduenko Aleksandr^{1*}

aduenko1@gmail.com

Strijov Vadim^{1,2}

strijov@ccas.ru

¹Moscow, Moscow Institute of Physics and Technology (State University)

²Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

The problem of multimodel selection when constructing models for two-class classification problem is considered. Multimodels are interpretable generalization of the single model case which addresses data inhomogeneity. Feature spaces may differ across the models. Moreover, a multimodel may contain a big number of similar models which leads to poor forecast quality and lack of interpretability. The method of statistical model comparison is suggested to address this problem. The multimodel is called adequate if the models constituting the multimodel are pairwise statistically distinguishable. The present authors suggest to introduce a similarity function between posterior distribution of model parameters for model comparison. Such a similarity function should be defined for the pair of distributions with different supports. Moreover, it should not distinguish the distributions one of which is non-informative. It is shown that Kulback–Leibler divergence, Jensen–Shannon, Hellinger, and Bhattacharaja distances do not satisfy this condition. The similarity function which satisfies the conditions is suggested. The asymptotic properties of similarity function distribution in case the model true parameters are identical have been obtained and proved. The usage of the suggested similarity function for model comparison is illustrated with synthetic data. Using the statistical properties of the introduced similarity function, the upper and the lower bounds on the maximum number of pairwise distinguishable models in a multimodel for a sample of the fixed size have been obtained. The lower bound is proved by providing the method of construction of adequate multimodel of the necessary size [1].

This research is funded by the Russian Foundation for Basic Research, grant 14-07-31205.

- [1] Aduenko, A. A., and Strijov V. V. 2014. Joint feature and object selection for multiclass classification of documents' collection. *Infocommunications Technologies* 1:47–53.

Вычислительно эффективные методы для обучения метрик и усиленные оценки риска метрических классификаторов

Иофина Галина Владимировна^{1*} giofina@mail.ru
Максимов Юрий Владимирович^{2,3,4} yurymaximov@iitp.ru
*Минаев Андрей Викторович*¹ andrew@minaev.net
*Поляков Юрий Сергеевич*¹ polyakovyury@gmail.com

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Институт проблем передачи информации РАН

³Москва, Московский физико-технический институт (ГУ),
Лаборатория структурных методов анализа данных
в предсказательном моделировании (ПреМоЛаб)

⁴Москва, Национальный исследовательский университет «Высшая
школа экономики»

Рассмотрена задача обучения метрик в задачах высокой размерности с большим числом классов. Основной мотивацией является усиление качества алгоритм k -ближайших соседей и доказательство усиленных оценок на риск классификаторов этого типа.

В реальных задачах распознавания возникает естественный выбор между сложностью (временем решения) задачи распознавания и качеством (точностью) решения задачи тем или иным алгоритмом. Основная идея предлагаемой конструкции состоит в том, что, учитывая размерность задачи, количество классов и внутреннюю структуру задачи, можно усилить известные оценки риска и скорость обучения классификатора для методов типа ближайших соседей.

В рассматриваемой конструкции задача построения оптимальной метрики сводится к задаче квадратичной оптимизации, для которой предлагается вычислительно эффективная процедура, скорость сходимости которой зависит от перечисленных параметров. Также представлен ряд экспериментов в задачах распознавания изображений [1].

Работа поддержана грантами РФФИ 14-07-31241 мол_а и 15-07-09121 а.

[1] *Iofina G., Maximov Yu., Minaev A., Polyakov Yu.* A new method for learning similarity functions in high dimensional problems // Pattern Recogn. Image Anal. 2015 (in press).

Computationally efficient metric learning with improved risk bounds

*Iofina Galina*¹★

giofina@mail.ru

Maximov Yury^{2,3,4}

yurymaximov@iitp.ru

*Minaev Andrey*¹

andrew@minaev.net

*Polyakov Yury*¹

polyakovyury@gmail.com

¹Moscow, Moscow Institute of Physics and Technology (State University)

²Moscow, Institute for Information Transmission Problems
of the Russian Academy of Sciences

³Moscow, Moscow Institute of Physics and Technology (State University),
Laboratory of Structural Methods in Predictive Modeling (PreMoLab)

⁴Moscow, National Research University “Higher School of Economics”

The problem of metric learning for high-dimensional multiclass classification problems is considered. The motivation of the research is to provide a k -nearest neighbor type algorithm that attains better risk estimates.

For real-life practical problems, there is a natural trade-off between computational complexity of learning and quality/accuracy of the algorithm. The main idea of the method is to take into account the number of classes, dimensionality, and internal structure of the problem.

In this paper, a novel algorithm is proposed for metric learning in k -nearest neighbor algorithm with improved risk guarantees. The algorithm leads to a quadratic optimization for which the authors propose an effective numerical procedure whose complexity depends on the parameters above. To support the algorithm, a series of experiments with image classification problems has been proposed [1].

The work is supported in part by the Russian Foundation for Basic Research, grant 14-07-31241 mol_a. The second author is also supported in part by the Laboratory of Structural Methods in Predictive Modeling Department of Control and Applied Mathematics, Moscow Institute of Physics and Technology, Russian Government, grant 11.G34.31.0073.

- [1] Iofina, G., Maximov Yu., Minaev A., and Polyakov Yu. 2015 (in press). A new method for learning similarity functions in high dimensional problems. *Pattern Recogn. Image Anal.*

Многоклассовая модель формы со скрытыми переменными

*Кириллов Александр Николаевич*¹ akirillov@cs.msu.ru
*Гавриков Михаил Игорьевич*² gavrmike@gmail.com
Лобачева Екатерина Максимовна^{3*} elobacheva@hse.ru
*Осокин Антон Александрович*⁴ anton.osokin@inria.fr
Ветров Дмитрий Петрович^{1,3} vetrovd@yandex.ru

¹Москва, Московский государственный университет
им. М. В. Ломоносова, Факультет ВМК

²Москва, ООО «Яндекс»

³Москва, Национальный исследовательский университет «Высшая
школа экономики», Факультет компьютерных наук

⁴Париж, SIERRA team, INRIA and École Normale Supérieure

Модель формы Больцмана (SBM) и ее многоклассовый вариант (MSBM) являются генеративными моделями, позволяющими задавать форму объекта на изображении. Многоклассовая модель обладает большей выразительной способностью, чем бинарная, но для ее обучения требуются данные с полной многоклассовой разметкой. В данной работе предложен новый алгоритм обучения MSBM, для применения которого достаточно неполной разметки данных, а именно: бинарной разметки и задания семян, указывающих приближенное расположение частей объектов. Семена при этом могут быть получены автоматически на основе использования детектора частей объектов. В теоретической части работы представлена графическая модель, задающая взаимосвязь между параметрами и переменными MSBM, бинарной и многоклассовой разметкой и семенами, а также описана основанная на ней EM-подобная процедура обучения MSBM по данным с неполной разметкой. На этапе экспериментов показано, что MSBM, обученная предложенным алгоритмом, превосходит по качеству SBM и сопоставима по качеству с MSBM, обученной на данных с полной разметкой.

Работа поддержана грантом РФФИ № 15-31-20596 и Microsoft Research в рамках проекта «Совместные исследования по компьютерному зрению в России».

- [1] *Кириллов А. Н., Гавриков М. И., Лобачева Е. М., Осокин А. А., Ветров Д. П.* Многоклассовая модель формы со скрытыми переменными // Интеллектуальные системы, 2015. Т. 19. Вып. 2. С. 75–95.

Deep part-based generative shape model with Latent variables

*Kirillov Alexander*¹

akirillov@cs.msu.ru

*Gavrikov Mikhail*²

gavrmike@gmail.com

*Lobacheva Ekaterina*³*

elobacheva@hse.ru

*Osokin Anton*⁴

anton.osokin@inria.fr

Vetrov Dmitry^{1,3}

vetrovd@yandex.ru

¹Moscow, Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics

²Moscow, Yandex

³Moscow, National Research University “Higher School of Economics”, Faculty of Computer Science

⁴Paris, SIERRA team, INRIA and École Normale Supérieure

The Shape Boltzmann Machine (SBM) and its multilabel version (MSBM) have been recently introduced as deep generative models that capture the variations of the object shape. While being more flexible, MSBM requires datasets with labeled parts of the objects for training. An algorithm for training MSBM using binary masks of objects and the seeds which approximately correspond to the locations of objects parts is presented. The latter can be obtained from part-based detectors in an unsupervised manner. A latent variable model and an EM-like (expectation–maximization) training procedure have been derived for adjusting the weights of MSBM using the deep learning framework. It is shown that the model trained by the suggested method outperforms SBM in the tasks related to binary shapes and is very close to the original MSBM in terms of the quality of multilabel shapes. Also, it was found out that the MSBM trained by new procedure significantly outperforms the MSBMs trained with multilabel segmentations obtained by some straight-forward heuristic procedure from the binary segmentations and the seeds.

This research is funded by the Russian Foundation for Basic Research, grant 15-31-20596, and by Microsoft Research, research initiative “Computer vision collaborative research in Russia.”

- [1] Kirillov, A. N., Gavrikov M., Lobacheva E., Osokin A., and Vetrov D. P. 2015. Deep part-based generative shape model with latent variables. *Intelligent Systems* 19(2):75–95.

Агрегация упорядоченных оценок в цветном шуме

Крымова Екатерина Александровна^{1,2*} krymova@phystech.edu

¹Москва, Институт проблем передачи информации РАН

²Москва, Московский физико-технический институт (ГУ)

Рассматривается задача восстановления функции регрессии по конечному числу наблюдений функции в гауссовском шуме, заданных в конечном числе детерминированных точек. Такая задача типична в непараметрической статистике и очень часто встречается на практике.

Предположим, что помимо наблюдений функции исследователю заранее известен фиксированный набор из конечного числа оценок функции. На основе этого набора оценок требуется построить новую оценку, качество которой было бы сравнимо с наилучшей (в смысле среднеквадратичного риска) оценкой из заданного множества. Под наилучшей оценкой из заданного множества понимаем оценку, риск которой минимален в заданном множестве оценок. Такую оценку в литературе часто называют оракулом, так как его значение исследователю неизвестно.

Получены новые оракульные неравенства (т. е. неравенства, которые связывают риск метода с риском оракула) для экспоненциальной агрегации упорядоченных оценок функции регрессии в предположении гетероскедастичного шума, а именно: шум предполагается коррелированным (ковариационная матрица известна) и дисперсия его различна в каждой точке наблюдения [1].

Работа поддержана грантом РФФИ № 15-07-09121.

- [1] *Krymova E.* Aggregation of ordered smoothers in colored noise // *J. Mach. Learn. Data Anal.*, 2015 (in press). Vol. 1. No.13. jmla.org/papers/doc/2015/JMLDA2015no13.pdf.

Aggregation of ordered smoothers in colored noise

Krymova Ekaterina^{1,2*}

krymova@phystech.edu

¹Moscow, Institute for Information Transmission Problems
of the Russian Academy of Sciences

²Moscow, Moscow Institute of Physics and Technology (State University)

The paper is devoted to the problem of recovery of one-dimensional functions given a set of noisy observations. Suppose that in addition, given a fixed set of a finite number of function estimates. Based on this set of estimates, it is necessary to construct a new estimator, the risk of which would be close to the risk of the “best” estimate (so-called oracle) in a given set. The present author proves new oracle inequalities for aggregation of regression function estimates in assumption of heteroscedastic Gaussian noise, namely, correlated Gaussian noise with different variances at each design point [1].

This research is funded by the Russian Foundation for Basic Research, grant 15-07-09121.

- [1] Krymova, E. 2015 (in press). Aggregation of ordered smoothers in colored noise. *J. Mach. Learn. Data Anal.* 1(13). jmla.org/papers/doc/2015/JMLDA2015no13.pdf

Комбинирование отношений порядка для восстановления предпочтения на наборе объектов

*Кузнецов Михаил Павлович** mikhail.kuznecov@phystech.edu
Москва, Московский физико-технический институт (ГУ)

Решается задача восстановления отношения предпочтения на наборе объектов. Задано множество D , состоящее из m объектов x_1, \dots, x_m . На множестве D задано n отношений порядка z_1, \dots, z_n : $z_j(x_i, x_k) = I(x_i \geq_j x_k)$. Задано целевое отношение порядка $z_0(x_i, x_k)$. Требуется по имеющимся описаниям объектов z_1, \dots, z_n восстановить полное отношение порядка, т. е. построить отображение $\varphi : x \mapsto y \in \mathbb{R}$, удовлетворяющее условию монотонности по всем z_1, \dots, z_n , и оптимальное восстанавливающее целевое отношение z_0 .

Для решения задачи определяется функция потерь, описывающая отклонение отношения порядка, задаваемого отображением φ , от целевого отношения z_0 . Ставится задача оптимизации функции потерь с ограничениями-неравенствами, задающими полидральный конус. Для получения явного решения используется представление конуса через его порождающие элементы. Показано, что при некоторых дополнительных ограничениях на параметры модели задача представляется в форме неотрицательной линейной регрессии.

Предлагается метод сокращения пространства параметров модели путем использования центральных точек рассматриваемых конусов. На основе метода разработан двухшаговый алгоритм восстановления отношения предпочтения на множестве объектов. Реализован вычислительный эксперимент, показывающий превосходство предложенного метода над базовыми [1].

Работа поддержана грантом РФФИ, проект офи_м_2013 № 13-01-12014.

- [1] *Kuznetsov M. P., Strijov V. V.* Methods of expert estimations concordance for integral quality estimation // Expert Syst. Appl., 2014. Vol. 41. No. 4. P. 1988–1996.

Combining order relations for estimating preferences over the set of objects

*Kuznetsov Mikhail**

mikhail.kuznecov@phystech.edu

Moscow, Moscow Institute of Physics and Technology (State University)

A problem of estimating preference relation over the set of objects $D = x_1, \dots, x_m$ is considered. The objects of D are described by the set of order relations z_1, \dots, z_n : $z_j(x_i, x_k) = I(x_i \geq_j x_k)$. Also, a target order relation $z_0(x_i, x_k)$ is given. The goal is to recover a full-order relation using the object descriptions z_1, \dots, z_n , that is, to construct a map $\varphi : x \mapsto y \in \mathbb{R}$ satisfying the monotonicity condition on z_1, \dots, z_n and optimally estimating the target relation z_0 .

To solve the problem, a loss function describing difference between the target relation z_0 and the relation defined by the map φ is introduced. The loss function generalizes the rank correlation notion. The preference estimation problem is formulated in terms of loss function minimization with inequality conditions, which define a polyhedral cone. To obtain an explicit solution, a finitely generated form of the cone is used. It is shown that by using some additional restrictions on the model parameters, the problem can be reformulated in terms of nonnegative least squares.

The author proposes to reduce model parameter space considering central points of the cones. On the basis of this method, a two-stage algorithm of the preference relation estimation over the set of objects is developed. The computations showed supremacy of the proposed method over the baseline [1].

The research is granted by the Russian Foundation for Basic Research, ofi_m_2013 13-01-12014.

- [1] Kuznetsov, M. P., and Strijov V. V. 2014. Methods of expert estimations concordance for integral quality estimation. *Expert Syst. Appl.* 41(4):1988–1996.

Многоклассовое распознавание в пространстве представлений с многоуровневым разрешением

Ланге Михаил Михайлович

lange_mm@ccas.ru

Ганебных Сергей Николаевич

sng@ccas.ru

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

Приведено обобщение многоклассовой модели распознавания образов по критерию ближайшего соседа [1] с учетом отказа от классификации. Рассматривается классификатор на основе схемы сравнения с эталонами в пространстве универсальных древовидных представлений образов с многоуровневым разрешением. На множестве представлений построено семейство мер различия образов и разделяющих функций, вычисляемых на последовательных уровнях разрешения. Разработана процедура обучения, включающая отбор эталонов и выбор пороговых значений разделяющих функций, используемых для реализации отказа.

Эффективность используемой процедуры обучения продемонстрирована ROC-кривыми (receiver operating characteristic) в терминах зависимости TPR (true positive rate) от FPR (false positive rate), характеристиками AUC (area under curve) и наименьшими значениями суммарной доли ошибок $((1-TPR)+FPR)/2$, для источников полутонных изображений лиц, подписей и жестов руки при различных уровнях разрешения.

Построен решающий алгоритм на основе параметрической стратегии поиска в сети эталонов с многоуровневым разрешением, реализующей иерархический и переборный поиск решения. Используя указанные источники изображений и процедуру скользящего контроля, выполнено тестирование классификаторов, построенных с отбором и без отбора эталонов. Для указанных классификаторов получены экспериментальные зависимости доли ошибок от вычислительной сложности решающего алгоритма при различных значениях параметра стратегии поиска.

Работа поддержана грантами РФФИ №№ 15-01-04671 и 15-07-09324.

- [1] Ланге М. М., Ганебных С. Н. Эффективность иерархической классификации в терминах соотношения качество–сложность // Машинное обучение и анализ данных, 2014. Т. 1. № 8. С. 1126–1136. jmla.org/papers/doc/2014/JMLDA2014no8.pdf.

Multiclass pattern recognition in a space of the multiresolution representations

Lange Mikhail

lange_mm@ccas.ru

*Ganebnykh Sergey**

sng@ccas.ru

Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

The paper contains a generalization of a multiclass pattern recognition model [1] which decision rule is added by rejection. A classifier that is based on comparing a submitted pattern with class selected templates in a space of the universal multiresolution object representations is developed. In a set of representations, dissimilarity measures and the measure-based discriminant functions in sequential resolution levels are defined. A learning procedure that consists in selecting the templates and estimating the rejection thresholds is developed. An efficiency of the used learning procedure is shown by ROC-curves (receiver operating characteristic) in terms of TPR (true positive rate) and FPR (false positive rate) relations as well as by AUC (area under curve) values and minimal error rates $((1-TPR)+FPR)/2$ that are obtained for sources of grayscale images of faces, signatures, and hand gestures taken in the sequential resolution levels. A decision algorithm that is based on a parametric strategy yielding hierarchical and exhaustive search for the decision in a multiresolution network of the templates is constructed. Using the above sources of the patterns, a cross-testing scheme is performed for two classifiers, particularly, with selecting and without selecting the templates. For these classifiers, experimental dependences of the error rate on a computational complexity of the decision algorithm are obtained by varying a parameter of the search strategy.

This research is funded by the Russian Foundation for Basic Research, grants 15-01-04671 and 15-07-09324.

- [1] Lange, M., and Ganebnykh S. 2014. An efficiency of hierarchical classification in terms of fidelity-complexity ratio. *Mach. Learn. Data Anal.* 1(8):1126–1136. jmla.org/papers/doc/2014/JMLDA2014no8.pdf.

Алгоритм приближенного поиска ближайшего цифрового массива на множестве пирамидальных представлений данных

*Ланге Михаил Михайлович**

lange_mm@ccas.ru

Ганебных Сергей Николаевич

sng@ccas.ru

Ланге Андрей Михайлович

lange_am@mail.ru

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

Предложен алгоритм приближенного поиска на множестве цифровых массивов, заданных m -мерными кубами из N^m элементов, ближайшего представителя к предъявляемому массиву по евклидовой мере. Известные алгоритмы поиска на множестве векторов евклидова пространства размерности $d \geq 1$ вектора, отстоящего от предъявляемого на расстояние не более $(1 + \varepsilon)D_{\min}$, где D_{\min} — расстояние до ближайшего представителя, а $\varepsilon > 0$ — допустимое отклонение, имеют сложность $O(c_{d,\varepsilon} \log n)$ при большой мощности n множества векторов и фиксированных d и ε . Константа $c_{d,\varepsilon}$ растет экспоненциально с увеличением d и полиномиально с уменьшением ε , что ограничивает применение таких алгоритмов для случая $d = N^m$ при $N > 10$ и $m > 1$ и, в частности, для изображений большого размера.

Предлагаемый алгоритм использует пирамидальные представления массивов с многоуровневым разрешением и иерархическую стратегию поиска на множестве представлений [1]. При фиксированном m и большом N получена асимптотическая оценка $O(N^{-m} \log N)$ доли вычислительной сложности иерархического алгоритма относительно сложности алгоритма полного перебора. Для множества изображений рукописных цифр размера 32×32 ($N = 32$, $m = 2$, $n = 50\,000$) в режиме скользящего контроля построена зависимость среднего значения величины ε от доли сложности иерархического алгоритма относительно сложности алгоритма перебора.

Работа поддержана грантами РФФИ № 15-01-04671 и № 15-07-07516.

- [1] *Lange M. M., Stepanov D. Yu.* Recognition of objects given by collections of multichannel images // *Pattern Recogn. Image Anal.*, 2014. Vol. 24. No. 3. P. 431–442.

Algorithm of approximate search for the nearest digital array in a set of pyramidal data representations

*Lange Mikhail**

lange_mm@ccas.ru

Ganebnykh Sergey

sng@ccas.ru

Lange Andrey

lange_am@mail.ru

Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

Given a set of m -dimensional cube-based digital arrays of N^m elements, an algorithm of approximate search in the set for the nearest neighbor of a submitted array is suggested. For the Euclidean vector space of dimension $d \geq 1$, the known algorithms of approximate nearest neighbor search for a vector with a distance not more than $(1 + \varepsilon)D_{\min}$ (D_{\min} is the distance between a submitted vector and its nearest neighbor and $\varepsilon > 0$ is the admissible value) possess a computational complexity $O(c_{d,\varepsilon} \log n)$ subject to a large value of the set cardinality n and fixed d and ε . Here, $c_{d,\varepsilon}$ grows exponentially in increasing d and polynomially in decreasing ε . So, this factor limits application of these algorithms in case of $d = N^m$ and $N > 10$, $m > 1$, and, particularly, for images of large size.

The developed algorithm uses pyramidal multiresolution representations of the arrays and a hierarchical search strategy in the given set of the representations [1]. In case of fixed m and large N , an asymptotic estimate $O(N^{-m} \log N)$ is given that yields a ratio of complexities of the hierarchical and exhaustive search algorithms. Using one time two fold cross-validation scheme for a set of hand written digit images of size 32×32 ($N = 32$, $m = 2$, and $n = 50\,000$), a mean value ε as a function of the ratio of the complexities of the hierarchical and exhaustive search algorithms is estimated by varying a parameter of the hierarchical search strategy.

This research is funded by the Russian Foundation for Basic Research, grants 15-01-04671 and 15-07-07516.

- [1] Lange, M. M., and Stepanov D. Yu. 2014. Recognition of objects Given by collections of multichannel images. *Pattern Recogn. Image Anal.* 24(3):431–442.

Многоклассовое распознавание в больших массивах данных

Маленичев Антон Александрович^{1*} malenichev@mail.ru

*Красоткина Ольга Вячеславовна*¹ o.v.krasotkina@yandex.ru

*Моттль Вадим Вячеславович*² vmottl@yandex.ru

¹Тула, Тульский государственный университет

²Москва, Вычислительный центр им. А. А. Дородницына РАН

Во многих областях технических и естественно-научных исследований часто возникает проблема многоклассовой классификации данных с учителем при условии, что размер обучающей совокупности достаточно велик. Имеющиеся в данный момент механизмы обучения либо обладают слишком большой вычислительной сложностью, чтобы производить обучение на больших массивах данных, либо принципиально способны работать лишь с двумя классами. Более того, в некоторых задачах требуется не только указывать самый предпочтительный класс нового поступившего объекта, но и указывать апостериорные вероятности его отнесения к тому или иному классу.

Идея работы заключается в построении механизма, опирающегося на хорошо известную задачу двухклассовой классификации с наложением на нее требования нечеткости решающего правила. Предлагается использовать метод многоклассового распознавания образов, основанный на методе двухклассовой классификации с нечетким решающим правилом. Такая постановка задачи способна уменьшить вычислительную сложность до сложности, близкой к линейной по количеству объектов. Также она выгодна тем, что существует возможность для распараллеливания вычислений на многопроцессорной машине или загрузкой исходной обучающей совокупности в память ЭВМ по частям, соответствующим каждой паре классов [1].

[1] *Malenichev A., Krasotkina O., Mottl V.* Multiclass learning in big data // J. Mach. Learn. Data Anal., 2015 (in press).

Multiclass learning in big data

*Malenichev Anton*¹*

malenichev@mail.ru

*Krasotkina Olga*²

o.v.krasotkina@yandex.ru

*Mottl Vadim*²

vmottl@yandex.ru

¹Tula, Tula State University

²Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

In many fields of engineering and scientific research, there exists a problem of supervised multiclass classification for big data. Currently available learning mechanisms have too big computational complexity to produce learning on big data or principally able to work with only two classes. Moreover, some problems require not only to indicate the most preferred class of new incoming object, but also indicate the posterior probabilities of its assignment to a particular class.

The focus of work is to construct a mechanism that relies on a well-known two-class classification task with adding to it the fuzzy decision rule requirements. So, the authors propose to use the method of multiclass pattern recognition based on two-class classification with fuzzy decision rule. This formulation of the problem can reduce the computational complexity close to linear by the number of objects. It is also advantageous because of the possibility for parallel computing on multiprocessor machine as well as in the ability of downloading the learning dataset in a computer memory by portions that correspond to each pair of classes [1].

[1] Malenichev, A., Krasotkina O., and Mottl V. 2015 (in press). Multiclass learning in big data. *J. Mach. Learn. Data Anal.*

Машина релевантных тегов

Молчанов Дмитрий Александрович^{1*} `dmolch111@gmail.com`
*Кондрашкин Дмитрий Андреевич*² `kondra2lp@gmail.com`
*Ветров Дмитрий Петрович*² `vetrovdyandex.ru`

¹Москва, Московский государственный университет
им. М. В. Ломоносова

²Москва, Национальный исследовательский университет «Высшая
школа экономики»

В задачах классификации и регрессии часто возникает проблема отбора релевантных признаков. Одним из подходов для решения этой проблемы является байесовский подход к выбору модели, основанный на максимизации так называемой обоснованности. Этот подход лежит в основе машины релевантных векторов, предложенной Типпингом в 2001 г. Машина релевантных векторов позволяет отбирать признаки в задачах регрессии и классификации, а в качестве априорного распределения на параметры модели используется нормальное распределение.

Авторы предлагают машину релевантных тегов для решения задачи бинарной классификации с бинарными признаками (тегами) и отбора релевантных тегов. Предложена новая вероятностная модель, где в качестве априорных распределений используются бета-распределения, которые, по мнению авторов, лучше подходят в случае бинарных признаков. Также предложен приближенный метод вывода в полученной модели, основанный на глобальных вариационных оценках. Экспериментально показано, что предложенная модель работает не хуже предыдущих подходов [1].

Работа поддержана грантом РФФИ № 15-31-20596-мол_a_вед, а также грантами Microsoft Research, research initiative: Computer vision collaborative research in Russia, Skoltech SDP Initiative, applications A1 and A2.

[1] *Молчанов Д. А., Кондрашкин Д. А., Ветров Д. П.* Машина релевантных тегов // Машинное обучение и анализ данных, 2015 (в печати). Т. 1. № 13. jmla.org/papers/doc/2015/JMLDA2015no13.pdf

Relevance tagging machine

Molchanov Dmitry^{1*}

dmolch111@gmail.com

*Kondrashkin Dmitry*²

kondra21p@gmail.com

*Vetrov Dmitry*²

vetrovd@yandex.ru

¹Moscow, Moscow State University

²Moscow, National Research University “Higher School of Economics”

The task of relevant feature selection arises in many classification and regression problems. One approach to address this task is Bayesian approach to model selection which is based on evidence maximization. Relevance vector machine (Tipping, 2001) is based on this approach. Relevance vector machine is successfully used to perform feature selection in classification and regression problems, it uses multivariate normal distribution as a prior to model parameters.

In this work, relevance tagging machine which is able to solve a binary classification problem with binary features (tags) is suggested and also, relevant feature selection is performed. A probabilistic model, in which beta distributions are used as priors, has been proposed that seems more reasonable than normal distributions in case of binary features. Also, an approximate inference algorithm based on global variational bounds has been suggested. An experimental investigation shows that the presented model is comparable to the previous state-of-the-art methods [1].

This research is funded by the Russian Foundation for Basic Research, grant #15-31-20596mol-a-ved, Microsoft Research, research initiative: Computer vision collaborative research in Russia, Skoltech SDP Initiative, applications A1 and A2.

- [1] Molchanov, D., Kondrashkin D., and Vetrov D. 2015 (in press). Relevance tagging machine. *J. Mach. Learn. Data Anal.* 1(13). jmla.org/papers/doc/2015/JMLDA2015no13.pdf

Оценка объема выборки в задачах классификации

*Мотренко Анастасия Петровна**

anastasiya.motrenko@phystech.edu

Москва, Московский физико-технический институт (ГУ)

Исследование направлено на решение проблемы выбора моделей при классификации выборок малой мощности. Для выбора оптимальной модели предлагается получить оценку необходимого объема выборки. Решение поставленной задачи актуально для анализа данных медицинских исследований. В частности, полное измерение иммунологических биомаркеров пациента является дорогостоящим, что влечет сложность сбора выборки достаточной мощности для принятия решений о классификации пациента.

Существующие методы оценки объема выборки не принимают в учет модель, используемую при решении задачи, что влечет за собой завышение получаемых оценок. Предлагается включить модель в процедуру оценивания объема выборки: объем выборки оценивается путем анализа функции распределения параметров модели. Параметры функции плотности распределения данных, оцениваемые по выборке, должны быть устойчивы относительно небольших возмущений состава выборки. Предлагается считать объем выборки m достаточным, если подвыборки данного объема m из рассматриваемого распределения будут отнесены некоторым критерием к одному распределению [1]. Для сравнения подвыборок вычисляется расстояние Кульбака–Лейблера между апостериорными распределениями параметров, оцененными на данных подвыборках, вводится статистический критерий принадлежности двух подвыборок одному распределению. Так как получаемые оценки зависят от выбранной модели классификации, предлагаемый метод оценки объема выборки позволяет оценить, является ли данный объем выборки адекватным применяемому методу классификации.

Работа поддержана грантом РФФИ № 15-37-50323.

- [1] *Motrenko A., Strijov V., Weber G.-W.* Bayesian sample size estimation for logistic regression // *J. Comput. Appl. Math.*, 2014. Vol. 255. P. 743–752.

Sample size estimation in classification problems

*Motrenko Anastasia**

anastasiya.motrenko@phystech.edu

Moscow, Moscow Institute of Physics and Technology (State University)

The aim of the study is to propose a model selection strategy in classification of small sample sets. The task is to estimate the sample size and select the optimal classification model on the basis of a small multivariate sample. Solving this problem is essential in analysis of biomedical data, such as immunological data, where costs of full examination of one patient can be high, which hampers obtaining sufficient samples for solving classification problem.

Existing methods of sample size estimation usually do not consider classification model, which leads to overestimation of sample size. It is proposed to include the model into the sample size estimation process via analysis of model parameters' probability distributions. To obtain adequate classification results, these distributions should be stable against subtle sample set variations. The author compares subsamples, computes Kullback–Leibler divergence between parameters' posteriori distributions, and declares sample size m sufficient, if subsamples of this size m are considered to pertain to the same probability distribution [1]. A statistical test for the two-sample problem, based on Kullback–Leibler divergence, has been introduced. Since the estimations, obtained by this way, depend on the chosen model, the proposed method allows to define if the current sample size is sufficient for a particular classification model.

This research is funded by the Russian Foundation for Basic Research, grant 15-37-50323.

- [1] Motrenko, A., Strijov V., and Weber G.-W. 2014. Bayesian sample size estimation for logistic regression. *J. Comput. Appl. Math.* 255:743–752.

К вопросу об эффективности бустинга в задаче классификации

*Неделько Виктор Михайлович**

nedelko@math.nsc.ru

Новосибирск, Институт математики им. С. Л. Соболева СО РАН

Исследуются причины высокой эффективности методов, основанных на композициях решающих функций, в частности бустинга. Показано, что одной из главных причин такой эффективности может быть использование эффекта независимости переменных. Для выявления особенностей метода исследуется его работа непосредственно на распределениях. Проводится сравнение аппроксимирующей способности бустинга и сплайнов. Также показана связь сложности композиции с достижимой величиной отступа.

В настоящее время использование методов построения решающих функций, основанных на композициях, является едва ли не необходимым условием для успешного решения практических задач. Тем не менее, до сих пор остается открытым вопрос: за счет чего композиции, в частности бустинг, зачастую превосходят другие методы?

В качестве обоснования (объяснения) эффективности бустинга часто приводится утверждение, что бустинг максимизирует отступ, а в силу известной оценки, чем больше отступ, тем меньше предполагаемая вероятность ошибки. Даже если согласиться с этим утверждением, следует заметить, что оно еще не объясняет, почему бустинг эффективен при решении практических задач, а фактически лишь заменяет этот вопрос вопросом, почему бустинг увеличивает отступ.

При этом, как оказывается, бустинг далеко не всегда увеличивает отступ. Эксперименты показывают, что в некоторых задачах распределение отступа с ростом композиции меняется не монотонно; более того, средний отступ (нормированный) может иметь тенденцию к уменьшению [1].

Работа поддержана грантом РФФИ № 14-01-00590.

- [1] *Неделько В. М.* К вопросу об эффективности бустинга в задаче классификации // Сибирский журнал чистой и прикладной математики, 2015. Т. 15. Вып. 2. С. 72–89.

On performance of boosting in classification problem

*Nedel'ko Victor**

nedelko@math.nsc.ru

Novosibirsk, Sobolev Institute of Mathematics of the Siberian Branch
of the Russian Academy of Sciences

The work provides some new explanation of effectiveness of the boosting methods. The main reason why boosting makes good decision functions on real world tasks is that the boosting utilizes some pattern of feature independence. Also, margin-based risk estimates are discussed with relation to boosting and it is shown that margin depends on complexity of composition.

Usually, the performance of boosting methods is explained via margin-based risk estimates: boosting typically increases margin, and the bigger margin the lower risk estimate. However, it was found that margin (after normalization) can decrease when composition grows.

Since margin maximization is not an immanent property of boosting, an effectiveness of boosting cannot be fully explained via margins theory only. As an important cause of boosting accuracy, feature independence pattern has been suggested [1].

This research is funded by the Russian Foundation for Basic Research, grant 14-01-00590.

- [1] Nedelko, V. 2015. On performance of boosting in classification problem. *Siberian J. Pure Appl. Math.* 15(2):72–89.

Тензорный поезд в марковском случайном поле

Новиков Александр Витальевич^{1,2} novikov@bayesgroup.ru

*Родоманов Антон Олегович*⁴ anton.rodomanov@gmail.com

*Осокин Антон Александрович*³ osokin@bayesgroup.ru

Ветров Дмитрий Петрович^{1,4}★ vetrovd@yandex.ru

¹Москва, Сколковский институт науки и технологий

²Москва, Институт вычислительной математики РАН

³Париж, INRIA

⁴Москва, Национальный исследовательский университет «Высшая школа экономики»

Вероятностные графические модели являются удобным инструментом для решения задач в таких областях, как компьютерное зрение, машинное обучение, анализ социальных сетей и др. Совместное распределение дискретных переменных можно задать через многомерный массив (тензор) ненормированной вероятности. В работе предлагается новый подход для работы с вероятностными графическими моделями, основанный на недавно предложенном разложении тензорного поезда (Tensor Train, TT). Разложение TT позволяет хранить тензор в компактном виде и эффективно применять к нему операции линейной алгебры. В данной работе предложен метод для перевода тензоров энергии и ненормированной вероятности марковского случайного поля в TT-формат. Свойства этого формата используются для подсчета нормировочной константы и поиска конфигурации наибольшей вероятности. Предоставлены теоретические гарантии точности оценки нормировочной константы. Предложенные методы сравниваются с аналогами на различных наборах данных [1].

Работа поддержана грантами РФФИ № 15-31-20596-мол_а_вед; Microsoft Research, research initiative: Computer vision collaborative research in Russia; Skoltech SDP Initiative, applications A1 and A2; результаты применения тензорного пакета поддержаны РФФИ № 14-11-00659.

- [1] *Novikov A., Rodomanov A., Osokin A., Vetrov D.* Putting MRFs on a tensor train // *J. Mach. Learn. Res.*, 2014. Vol. 32. P. 811–819. jmlr.org/proceedings/papers/v32/novikov14.pdf.

Putting Markov random fields on a tensor train

Novikov Alexander^{1,2}

novikov@bayesgroup.ru

*Rodomanov Anton*⁴

anton.rodomanov@gmail.com

*Osokin Anton*³

osokin@bayesgroup.ru

Vetrov Dmitry^{1,4}★

vetrovd@yandex.ru

¹Moscow, Skolkovo Institute of Science and Technology

²Moscow, Institute of Numerical Mathematics of the Russian Academy of Sciences

³Paris, INRIA

⁴Moscow, National Research University “Higher School of Economics”

Markov random fields (MRF) have become a popular tool for many applications in the fields of computer vision, machine learning, social networking, etc. One way of defining the joint distribution of discrete variables is to express it as a multidimensional array (tensor) of (unnormalized) probabilities. A new framework is presented for dealing with probabilistic graphical models. The suggested approach relies on the recently proposed Tensor Train (TT) format of a tensor that while being compact allows for efficient application of linear algebra operations. A way to convert the energy of an MRF to the TT-format is presented and it is shown how one can exploit the properties of the TT-format to attack the tasks of the partition function estimation and the MAP (maximum *a posteriori* probability) inference. Theoretical guarantees on the accuracy of the proposed algorithm for estimating the partition function have been provided and the suggested methods have been compared against several state-of-the-art algorithms [1].

This research is funded by the Russian Foundation for Basic Research, grant No. 15-31-20596mol-a-ved; by Microsoft Research, research initiative: Computer vision collaborative research in Russia; by Skoltech SDP Initiative, applications A1 and A2; the results of the tensor toolbox application are supported by Russian Science Foundation No. 14-11-00659.

- [1] Novikov, A., Rodomanov A., Osokin A., and Vetrov D. 2014. Putting MRFs on a tensor train. *J. Mach. Learn. Res.* 32:811–819. jmlr.org/proceedings/papers/v32/novikov14.pdf.

Использование обобщенных прецедентов для сжатия больших выборок при обучении

*Рязанов Владимир Васильевич*¹

rvvccas@mail.ru

*Виноградов Александр Петрович*¹

vngrccas@mail.ru

*Лаптин Юрий Петрович*²★

laptin_yu_p@mail.ru

¹Москва, Вычислительный центр им. А. А. Дородницына РАН

²Киев, Институт кибернетики НАН Украины

Анализируется роль внутренне присущих и привносимых структур данных в построении эффективных алгоритмов распознавания. Исследуются понятие обобщенного прецедента и основанные на его использовании методы сокращения размерности задач. Предложены два новых подхода к проблеме, основанные на позиционном представлении данных и на использовании групповых средних для элементарных логических закономерностей.

Показано, что в случае позиционного представления пространства признаков R^N может быть редуцировано до двумерного пространства гиперкубов позиционной иерархии. Редуцированное представление представляет собой одномерную развертку R^N , нагруженную весами обобщенных прецедентов. Разработан и исследован новый метод сжатия обучающей выборки, основанный на использовании групповых средних по кластерам элементарных логических закономерностей и на их использовании в качестве обобщенных прецедентов в преобразованном $(N + 1)$ -мерном признаковом пространстве.

Вычислительный эксперимент проводился для нескольких типов обобщенных прецедентов на реальных задачах. Хорошие предварительные результаты подтверждают наличие здесь новых возможностей и перспективность использования обобщенных прецедентов в задачах распознавания с большими выборками обучающих данных [1].

Работа поддержана грантами РФФИ № 15-01-05776-а и № 14-01-90413 Укр_а.

- [1] *Ryazanov V. V.* Using generalized precedents for big data sample compression at learning // J. Mach. Learn. Data Anal., 2015 (in press). Vol. 1. No. 14. jmla.org/papers/doc/2015/JMLDA2015no14.pdf.

Using generalized precedents for big data sample compression at learning

*Ryazanov Vladimir*¹

rvvccas@mail.ru

*Vinogradov Alexander*¹

vngrccas@mail.ru

Lapin Yuriy^{2*}

lapin_yu_p@mail.ru

¹Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

²Kiev, Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine

The role of intrinsic and introduced data structures in constructing efficient recognition algorithms is analyzed. The concept of generalized precedent and based on its use methods of reduction of the dimension of tasks have been investigated. Two new approaches to the problem based on positional data representation and on cluster means for elementary logical regularities are proposed.

It was shown that in case of positional data representation, the feature space R^N can be reduced to two-dimensional space of hypercubes of positional hierarchy. Reduced representation realizes the one-dimensional scan of R^N , which is loaded with weights of generalized precedents. A new method of training data compression has been developed and investigated based on the use of cluster means for elementary logical regularities and on its use as generalized precedents in transformed $(N + 1)$ -dimensional feature space.

Computational experiment was made for several types of generalized precedents on real tasks. Good preliminary results approve the new opportunities and open prospects of the use of generalized precedents in recognition tasks with big data samples [1].

This research is funded by the Russian Foundation for Basic Research, grants 15-01-05776-a and 14-01-90413 Ukr_a.

- [1] Ryazanov, V. 2015 (in press). Using generalized precedents for big data sample compression at learning. *J. Mach. Learn. Data Anal.* 1(14). jmla.org/papers/doc/2015/JMLDA2015no14.pdf.

Оценка эффекта множественного тестирования при поиске закономерностей в данных высокой размерности

Сенько Олег Валентинович^{1*}

senkoov@mail.ru

*Морозов Алексей Михайлович*²

alxmoro3ov@gmail.com

*Кузнецова Анна Викторовна*³

azfor@narod.ru

*Клименко Людмила Леонидовна*⁴

klimenkoll@mail.ru

¹Москва, Федеральный исследовательский центр «Информатика и управление» РАН

²Москва, Московский государственный университет им. М. В. Ломоносова

³Москва, Институт биохимической физики РАН

⁴Москва, Институт химической физики РАН

Разработка методов поиска статистически достоверных эмпирических закономерностей является одной из приоритетных задач интеллектуального анализа данных. Одной из возможных технологий поиска таких закономерностей является метод оптимальных достоверных разбиений, который использует для статистической верификации перестановочный тест. В условиях высокой размерности данных оценка достоверности двумерных закономерностей существенно осложняется проблемой множественного тестирования. Использование стандартного метода коррекции Бонферрони требует фиксации чрезвычайно жестких и практически редко достижимых порогов при отборе достоверных закономерностей при размерности данных выше 100. Серия Монте-Карло экспериментов была проведена для оценки истинной достоверности закономерностей, выявленных при решении биомедицинской задачи изучения связи уровня фактора роста сосудов с широким набором биологических показателей. Набор закономерностей, найденных в исходной выборке, сравнивался с наборами закономерностей, найденных в 50 случайных выборках, полученных из исходной путем случайных перестановок значений целевой переменной. Эксперименты показали, что многие закономерности, не соответствующие критерию Бонферрони, на самом деле оказываются достоверными [1].

- [1] *Senko O., Kostomarova I., Kuznetsova A.* Modification of the method of optimal valid partitioning for comparison of patterns related to the occurrence of ischemic stroke in two groups of patients // *Pattern Recogn. Image Anal.*, 2014. Vol. 24. Iss. 1. P. 114–123.

Evaluating of multiple testing effect in task of regularities search in multidimensional data.

Senko Oleg^{1*}

senkoov@mail.ru

*Morozov Aleksey*³

alxmopo3ov@gmail.com

*Kuznetsova Anna*³

azfor@narod.ru

*Klimenko Lyudmila*⁴

klimenkoll@mail.ru

¹Moscow, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences

²Moscow, Lomonosov Moscow State University

³Moscow, Emanuel Institute of Biochemical Physics of the Russian Academy of Sciences

⁴Moscow, Semenov Institute of Chemical Physics of the Russian Academy of Sciences

Development of method for statistically valid regularities discovery is one of the most important data mining problems. One of the possible techniques of regularities search is the method of optimal valid partitioning, using permutation test for statistical verification. In high-dimensional tasks, verification task becomes more complicated due to the problem of multiple testing. Standard Bonferroni correction is based on very strong validity thresholds that rarely are practically achievable when dimension is greater 100. Set of Monte-Carlo experiments was done to evaluate true validity of found regularities in the following biomedical task: study of relationship between vessels growth factor levels and wide set of biological and biochemical indicators, microelements levels, and electrical potentials of head. Set of regularities found in initial data set was compared with the sets of regularities that were found in 50 random data sets obtained from the original by random permutations of the target variable positions. Experiments have shown that many of regularities that were rejected according to the Bonferroni rule actually were valid. The experiments allow to evaluate true validity conditions [1].

- [1] Senko, O., Kostomarova I., and Kuznetsova A. 2014. Modification of the method of optimal valid partitioning for comparison of patterns related to the occurrence of ischemic stroke in two groups of patients. *Pattern Recogn. Image Anal.* 24(1):114–123.

Об основах метрического анализа плохо формализованных задач распознавания и классификации

Торшин Иван Юрьевич^{1*}

tiy135@yahoo.com

Рудаков Константин Владимирович^{1,2}

rudakov@ccas.ru

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Вычислительный центр им. А. А. Дородницына РАН

Во многих областях современной науки существуют задачи, адекватная формализация которых принципиально необходима для получения теоретически и практически важных результатов. В рамках терминологии научной школы академика РАН Ю. И. Журавлёва, формализованными являются задачи с однозначно задаваемой матрицей информации и информационной матрицей. В настоящей работе рассмотрен комплекс проблем, связанных с формализацией задач распознавания/классификации и сформулированы основы формализма для проведения метрического анализа плохо формализованных задач. Процесс формализации задачи может быть представлен как последовательный переход от множества исходных описаний к определенной топологии, затем к решетке и потом к метрическому пространству. Показано, что свойство регулярности задач по Журавлёву достаточно для существования биективных отображений между рассматриваемыми конструкциями. Рассмотрены возможности приложения разрабатываемого аппарата к таким важным проблемам формализации задач, как введение метрик на множествах признаков описаний, метрик на множествах объектов и анализ «взаимодействий» между разнородными признаковыми описаниями.

Работа поддержана грантом РФФИ № 12-07-00485.

- [1] *Torshin I. Yu., Rudakov K. V.* On the theoretical basis of the metric analysis of poorly formalized problems of recognition and classification // *Pattern Recogn. Image Anal.*, 2015 (в печати).

Metric analysis of poorly formalized problems

*Torshin Ivan*¹★

tiy135@yahoo.com

Rudakov Konstantin^{1,2}

rudakov@ccas.ru

¹Moscow, Moscow Institute of Physics and Technology (State University)

²Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

In many areas of modern science, there are problems adequate formalization of which is indispensable for obtaining both practically and theoretically important results. In terms of the scientific school of Yu. I. Zhuravlev, a formalized problem is uniquely defined by the matrix of information and the information matrix. Here, the complex of issues related to the formalization of the problems of pattern recognition/classification is considered and basics of a universal formalism for metric analysis of poorly formalized problems is proposed. Thus, formalization of the problem can be represented as a gradual transition from the set of original descriptions to a particular topology, then, to a lattice, and after that, to a certain metric space. It is shown that the property of Zhuravlev's regularity is sufficient for the existence of bijective mappings between these mathematical constructs. The possibilities of application of the apparatus developed are illustrated on several issues important for the formalization of the problems: introduction of metrics on the sets of the features, on the sets of the objects, and analysis of "interactions" between dissimilar feature descriptions.

This research is in part funded by the Russian Foundation for Basic Research, grant 12-07-00485.

- [1] Torshin, I. Yu., and Rudakov K. V. 2015 (in press). On the theoretical basis of the metric analysis of poorly formalized problems of recognition and classification. *Pattern Recogn. Image Anal.*

Бустинг приближенных алгоритмов в приложении к полиэдральной отделимости комитетного типа

*Хачай Михаил Юрьевич**

`mkhachay@imm.uran.ru`

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

Комбинаторная оптимизация и теория обучения — две тесно взаимодействующие и активно развивающиеся области современной теоретической информатики. Традиционно направление взаимодействия между этими разделами было связано с использованием точных и приближенных методов решения задач комбинаторной оптимизации для построения эффективных (суб)оптимальных процедур обучения, схем кластеризации, отбора признаков и т. п. Однако публикации последних лет демонстрируют возможность обратного взаимодействия, например, связанного с использованием процедур обучения при построении эффективных точных или приближенных алгоритмов с гарантированными оценками и полиномиальных приближенных схем для задач комбинаторной оптимизации.

Приводится пример применения схемы бустинга разработки приближенных алгоритмов с рекордными оценками точности для задачи о минимальном аффинном разделяющем комитете (MASC) при дополнительном условии фиксированной размерности пространства $n > 1$ и общности положения разделяемых множеств. Предлагаемый в работе игровой подход является развитием результатов Й. Фройнда и С. Ароры [1].

Работа поддержана грантом РФФИ № 14-11-00109.

- [1] *Khachay M.* Committee polyhedral separability: Complexity and polynomial approximation // *Mach. Learn.*, 2015. <http://dx.doi.org/10.1007/s10994-015-5505-0>.

Boosting of polynomial time approximation algorithms for committee polyhedral separability

*Khachay Michael**

mkhachay@imm.uran.ru

Ekaterinburg, Krasovsky Institute of Mathematics and Mechanics
of the Ural Branch of the Russian Academy of Sciences

The Minimum Affine Separating Committee (MASC) combinatorial optimization problem, which is related to ensemble machine learning techniques on the class of linear weak classifiers combined by the rule of simple majority, has been considered. Actually, the MASC problem is a mathematical formalization of the famous Vapnik–Chervonenkis (VC) principle of structural risk minimization in the mentioned class of classifiers. According to this principle, it is required to construct the best performance ensemble classifier belonging to a family of the least possible VC-dimension.

It is known that the MASC problem is NP-hard and remains intractable in spaces of any fixed dimension $n > 1$ even under an additional constraint on the separated sets to be in general position. This special case of the MASC problem called MASC-GP(n) is the main subject of interest of the present paper.

To design polynomial-time approximation algorithms for a class of combinatorial optimization problems containing the MASC problem, a new framework has been suggested, adjusting the well-known Multiple Weights Update method. Following this approach, polynomial-time approximation algorithms with the state-of-the-art approximation guarantee for the MASC-GP(n) problem has been constructed. The results obtained provide a theoretical framework for learning a high-performance ensembles of affine classifiers [1].

This research is supported by the Russian Science Foundation, grant 14-11-00109.

- [1] Khachay, M. 2015. Committee polyhedral separability: Complexity and polynomial approximation. *Mach. Learn.* Available at: <http://dx.doi.org/10.1007/s10994-015-5505-0> (accessed July 30, 2015).

Построение монотонного классификатора для задач медицинской диагностики

Швец Михаил Юрьевич^{1*} mikhail.shvets@phystech.edu

*Зухба Анастасия Викторовна*¹ a__l@mail.ru

Воронцов Константин Вячеславович^{1,2} voron@forecsys.ru

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Вычислительный центр им. А. А. Дородницына РАН

Линейные векторные модели классификации хорошо зарекомендовали себя во многих прикладных задачах. Расширение линейной модели до семейства монотонных нелинейных функций повышает точность аппроксимации обучающей выборки, но увеличивает риск переобучения. Для сокращения сложности модели ставится задача выбора минимального информативного набора признаков при одновременной монотонизации выборки путем отбора максимального подмножества опорных объектов. Предлагаются вычислительно эффективные эвристические алгоритмы для решения NP-трудной задачи одновременного отбора объектов и признаков при построении монотонного классификатора. Для отбора признаков предлагаются жадные алгоритмы, основанные на предварительном ранжировании признаков по их информативности в линейном наивном байесовском классификаторе.

Эксперименты проводились на задаче диагностики 14 заболеваний по электрокардиограмме длительностью 5–8 мин. В качестве признаков использовались частоты появления 216 паттернов variability сердечного ритма, охватывающих 4 соседних кардиоцикла. Для каждого заболевания найдены информативные подпространства признаков низкой размерности, подобраны сочетания эвристик в алгоритмах одновременного отбора объектов и признаков, исследована возможность повышения качества классификации с помощью линейной композиции монотонных классификаторов. Получены достаточно высокие показатели чувствительности и специфичности (90% и выше) [1].

Работа поддержана РФФИ, грант № 14-07-31240.

- [1] *Швец М. Ю., Зухба А. В.* Монотонные классификаторы для задач медицинской диагностики // Математическая биология и биоинформатика, 2015 (в печати). www.matbio.org.

Learning isotonic classifier for medical diagnostics

*Shvets Mikhail*¹*

mikhail.shvets@phystech.edu

*Zukhba Anastasia*¹

a_l@mail.ru

Vorontsov Konstantin^{1,2}

voron@forecsys.ru

¹Moscow, Moscow Institute of Physics and Technology (State University)

²Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

Linear vector models for classification are successfully used in a wide range of applications. Extending a linear model to the family of isotonic nonlinear functions improves the accuracy of the training set approximation but increases the risk of overfitting. Feature selection helps to reduce overfitting by determining an optimal complexity of the model. On the other hand, prototype selection helps to monotone the training set by selecting the largest subset of objects consistent with the monotonicity constraint. The problem of simultaneous feature and prototype selection for learning isotonic classifier is NP-hard. In this paper, computationally efficient heuristic algorithms is suggested to solve it. The authors propose greedy feature selection algorithms that perform preliminary ranking of features based on their weights in the Naïve Bayes linear classifier.

The experiments were carried out for medical diagnostics of 14 diseases by 5–8-minute electrocardiogram records. The features were based on the frequencies of 216 patterns of heart rate variability, covering 4 neighboring cardiocycles. As a result, the informative feature subspaces of low dimensionality for each disease were found, the combinations of suitable heuristics for the algorithms of simultaneous feature and prototype selection were adjusted, and the improvement of classification by using linear composition of monotonic classifiers were investigated. The authors obtained high sensitivity and specificity values, 90% and higher depending on the disease [1].

This research is funded by the Russian Foundation for Basic Research, grant 14-07-31240.

- [1] Shvets, M., and Zukhba A. 2015 (in press). Isotonic classifier for medical diagnostics *Math. Biol. Bioinf.* www.matbio.org.

Агрегирующие корректные операции над алгоритмами

Шибзухов Заур Мухадинович^{1,2*}

szport@gmail.com

¹Нальчик, Институт прикладной математики и автоматизации

²Москва, Центр «Антистихия» МЧС России

Обсуждается класс агрегированно корректных алгоритмов распознавания и прогнозирования и агрегирующие корректные операции над ними, которые произвольный конечный набор агрегированно корректных алгоритмов переводят в агрегированно корректный алгоритм. Такие операции расширяют классы базовых алгоритмов при помощи агрегирующих операций определенного вида и позволяют строить новые композиции алгоритмов распознавания и прогнозирования.

Вводятся понятия выпуклой функции и агрегирующей функции относительно пары идемпотентных агрегирующих операций и на их основе строятся достаточные условия того, что идемпотентная агрегирующая операция над алгоритмами является агрегированно корректной относительно заданного агрегированного функционала качества алгоритмов.

Полученные результаты обосновывают новые способы построения операций над алгоритмами, которые сохраняют свойство агрегированной корректности алгоритмов [1].

Работа выполнена при поддержке гранта РФФИ №15-01-03381а.

[1] *Шибзухов З. М.* Агрегирующие корректные операции над алгоритмами // Докл. РАН, 2015. Т. 462. № 6. С. 649–652.

Aggregation correct operations on algorithms

Shibzukhov Zaur^{1,2*}

szport@gmail.com

¹Nalchick, Institute of Applied Mathematics and Automation

²Moscow, Center “Antistikhiya” of Emercom of Russia

The paper discusses the classes of aggregationally correct algorithms for recognition and forecasting and aggregation operations on algorithms that an arbitrary finite set of aggregationally correct algorithms transform into aggregationally correct algorithm. Such operations expand basic classes of algorithms using a certain type of aggregation operations and allow to construct new compositions of algorithms for recognition and forecasting.

The concepts of a convex function and aggregate function with respect to a pair of idempotent aggregation operations are introduced. On the base of these concepts, sufficient conditions have been introduced that idempotent aggregating operation on algorithms is an aggregationally correct with respect to a specified aggregation functional of algorithm’s quality.

The results justify new methods of construction operations on algorithms that preserve the property of aggregated correctness of algorithms [1].

This research is funded by the Russian Foundation for Basic Research, grant No. 15-01-03381a.

[1] Shibzukhov, Z. M., 2015. Aggregation correct operations on algorithms. *Dokl. Math.* 91(3):391–393.

О моделях нейронов агрегирующего типа

Шибзухов Заур Мухадинович^{1,2} szport@gmail.com
Чередников Денис Юрьевич^{2*} denis.cherednikov@gmail.com

¹Нальчик, Институт прикладной математики и автоматизации

²Москва, Московский педагогический государственный университет

Описывается новый класс моделей искусственных нейронов агрегирующего типа, которые включают в себя классические функциональные модели, в том числе и со сложными синапсами. Такие модели основаны на применении агрегирующих функций для вычисления суммарного потенциала нейрона, вкладов в него сложных синапсов и синаптических кластеров.

Одной из таких моделей является модель $\Sigma\Pi$ -нейрона. Ключевым преимуществом данной модели является тот факт, что $\Sigma\Pi$ -нейрон более адекватно соответствует процессам, протекающим в коре головного мозга. Доказано, что $\Sigma\Pi$ -нейроны вполне при определенных условиях способны представлять произвольные дискретные функции, определенные на конечных множествах. Модель $\Sigma\Pi$ -нейрона естественным образом расширяется до модели агрегирующего нейрона, в которой вместо суммы используется квазиаддитивная агрегирующая функция, а вместо произведения используется квазимультимпликативная агрегирующая функция. Доказано, что такая модель агрегирующего нейрона также способна представлять произвольные дискретные функции, определенные на конечных множествах. Такие нейроны строятся при помощи конструктивной процедуры обучения. При этом, из-за неоднозначности выбора допустимых квазимультимпликативных элементов, можно построить семейство таких нейронов, которые все являются корректными на обучающем множестве. Конструктивность процесса обучения позволяет также минимизировать сложность сложных синапсов и синаптических кластеров [1].

Работа выполнена при поддержке гранта РФФИ №15-01-03381а.

- [1] Шибзухов З. М., Чередников Д. Ю. О моделях нейронов агрегирующего типа // Машинное обучение и анализ данных, 2015. Т. 1. № 12. С. 1706–1716. jmla.org/papers/doc/2015/JMLDA2015no12.pdf

On the models of neurons of aggregation type

Shibzukhov Zaur^{1,2}

szport@gmail.com

Cherednikov Denis^{2*}

denis.cherednikov@gmail.com

¹Nalchik, Institute of Applied Mathematics and Automatization

²Moscow, Moscow Pedagogical State University

A new class of models of artificial neurons is described, which includes classical functional models with complex synapses. Such models are based on application of aggregate functions for evaluation of total neuron potential and inclusions of complex synapses and synaptical clusters.

One of these models is $\Sigma\Pi$ -neuron. The main advantage of such model is that $\Sigma\Pi$ -neuron adequately corresponds to the information processing occurring in the cerebral cortex. It has been proven that $\Sigma\Pi$ -neurons within certain limits are able to present arbitrary discrete functions, determined in finite sets. $\Sigma\Pi$ -neuron model can be expanded naturally to aggregate neuron model in which instead of addition, a quasi-additive aggregate function is used, and instead of multiplication, a quasi-multiplicative aggregate function is used. It has been proven that such aggregate neuron model is also able to present arbitrary discrete functions, determined in finite sets. Such neurons can be built by constructive training process. At the same time, because of the ambiguity of the selection range of quasi-multiplicative properties of the components, a set of such neurons can be constructed, which are all correct in the training set. Also, constructive learning process allows to minimize the complicacy of complex synapses and synaptic clusters [1].

This research is funded by Russian Foundation for Basic Research, grant 15-01-03381a.

- [1] Shibzukhov, Z. M., and Cherednikov D. Y. 2015. On the models of neurons of aggregation type. *J. Mach. Learn. Data Anal.* 1(12):1706–1716. jmla.org/papers/doc/2015/JMLDA2015no12.pdf

Коллективный кластерный анализ с весами, основанный на коассоциативных матрицах

*Бериков Владимир Борисович**

berikov@math.nsc.ru

¹Новосибирск, Институт математики им. С. Л. Соболева СО РАН

Исследуется метод коллективного кластерного анализа, основанный на построении коассоциативной матрицы различий, с учетом весов базовых алгоритмов, которые вычисляются с использованием индексов качества группировки. Вводится вероятностная модель ансамблевой классификации, исследуются ее свойства.

Существует несколько основных способов построения коллективных решений кластерного анализа. В рамках первого направления от ансамбля требуют консенсуса, т. е. некоторой наилучшей степени согласованности с результатами отдельных алгоритмов. Другое направление основано на вычислении коассоциативной матрицы, определяющей, как часто пара объектов оказывается в одном и том же кластере в разных вариантах разбиения.

В данной работе развивается направление, основанное на коассоциативных матрицах, с учетом весов алгоритмов. При этом в итоговом коллективном решении предлагается учитывать степень «компетентности» (вес) каждого алгоритма на разных парах объектов. Для обоснования критерия качества ансамбля с равными весами ранее использовалась модель коллективного кластерного анализа, впервые введенная в работе [1]. В работе [2] была описана модификация данной модели, учитывающая веса различных алгоритмов.

В предлагаемой работе проводится дальнейшее развитие модели: дополнительно учитываются значения индексов качества группировки, что позволяет более полно отразить характеристики алгоритмов, включенных в ансамбль.

Работа поддержана грантом РФФИ №14-07-00249.

- [1] *Berikov V.* A latent variable pairwise classification model of a clustering ensemble // Multiple classifier systems / Eds. C. Sansone, J. Kittler, F. Roli. Lecture notes on computer science ser. — Heidelberg: Springer, 2011. Vol. 6713. P. 279–288.
- [2] *Berikov V.* Weighted ensemble of algorithms for complex data clustering // Pattern Recognition Letters, 2014. Vol. 38. P. 99–106.

Ensemble clustering using weights based on coassociation matrices

*Berikov Vladimir**

berikov@math.nsc.ru

Novosibirsk, Sobolev Institute of Mathematics of the Siberian Branch
of the Russian Academy of Sciences

The author investigates the method of ensemble clustering based on weighted averaged coassociation matrix, where the weights of ensemble elements are determined with the use of cluster validity indices. Probabilistic model of ensemble clustering has been introduced and theoretically investigated.

A number of approaches for finding ensemble clustering decision were suggested in literature. In consensus-style procedures, it is required to reach optimal degree of consistency with the results of individual algorithms. Another approach is based on the notion of averaged coassociation matrix. This matrix defines how often object pairs fall into different (or the same) clusters over all clustering variants.

The present paper follows the ensemble clustering approach based on coassociation matrices. For the formation of final collective decisions, weights of algorithms on different object pairs was considered. In [1], a probabilistic model of clustering ensemble with equal weights was suggested. The model was developed in [2] for analysis of ensembles composed of different clustering algorithms. The present paper further elaborates the ideas of these works: for ensemble variants, the author additionally introduces weights, which are dependent on cluster validity indices.

This research was funded by the Russian Foundation for Basic Research, grant 14-07-00249.

- [1] Berikov, V. 2011. A latent variable pairwise classification model of a clustering ensemble. *Multiple classifier systems*. Eds. C. Sansone, J. Kittler, and F. Roli. Lecture notes on computer science ser. Heidelberg: Springer. 6713:279–288.
- [2] Berikov, V. 2014. Weighted ensemble of algorithms for complex data clustering. *Pattern Recognition Letters* 38:99–106.

Цензурирование ошибочно классифицированных объектов выборки

Борисова Ирина Артемовна^{1,2,3*}

biamia@mail.ru

Кутненко Ольга Андреевна^{1,2,3}

olga@math.nsc.ru

¹Новосибирск, Институт математики им. С. Л. Соболева СО РАН

²Новосибирск, Новосибирский государственный университет

³Новосибирск, Конструкторско-технологический институт
вычислительной техники СО РАН

Задача цензурирования зашумленной выборки является одной из центральных проблем анализа данных. Под зашумленной выборкой будем понимать выборку, изначально содержащую неверно классифицированные объекты. В работе описан алгоритм цензурирования таких данных, ориентированный только на локальные характеристики объектов выборки.

Для оценки вероятности принадлежности объекта z к одному из двух образов A или B используется тернарная относительная мера — функция конкурентного сходства (FRIS-функция), оценивающая сходство объекта z с множеством A в конкуренции с множеством B . В фиксированном признаковом пространстве цензурирование состоит в последовательном удалении объектов, максимально ухудшающих качество описания выборки (или оценку разделимости классов), которая зависит от конкурентного сходства объектов со своим образом в конкуренции с другим образом и от числа рассматриваемых объектов. Результаты тестирования предложенного алгоритма на широком спектре модельных задач позволили сделать вывод, что объекты, удаленные до точки перегиба функции, описывающей разделимость классов, как правило, являются выбросами, искажающими структуру данных. Поэтому исключение таких объектов из анализируемой выборки повышает надежность распознавания [1].

Работа поддержана грантом РФФИ № 14-01-00039.

- [1] *Борисова И. А., Кутненко О. А.* Цензурирование ошибочно классифицированных объектов выборки // Машинное обучение и анализ данных, 2015. Т. 1. № 11. С. 1632–1641. jmla.org/papers/doc/2015/no11/Borisova2015Outliers.pdf.

Outliers detection in datasets with misclassified objects

Borisova Irina^{1,2,3,*}

biamia@mail.ru

Kutnenko Olga^{1,2,3}

olga@math.nsc.ru

¹Novosibirsk, Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences

²Novosibirsk, Novosibirsk State University

³Novosibirsk, Design Technological Institute of Digital Techniques of the Siberian Branch of the Russian Academy of Sciences

The problem of outliers detection is one of the important problems in Data Mining. Here, outliers are considered as initially misclassified objects of the dataset. This paper describes an algorithm of censoring such data, which focuses only on the local characteristics of objects of the dataset.

To calculate the level of belonging of the object z to one of two classes A and B , the ternary relative measure called the function of rival similarity (FRiS-function) is used. It evaluates the similarity of the object z with the class A in competition with the class B . Censoring procedure in a fixed feature space consists of sequential removals of objects, which deteriorate the quality of dataset description (a value of classes' separability) in a strongest way. This value depends on the number of objects in the dataset and similarity of objects with their class in competition with the rival class. The proposed algorithm was tested on a wide range of model problems. According to the results obtained, it is possible to conclude that the objects, which were deleted before the inflection point of the classes separability function, usually distort the structure of the data. Therefore, their exclusion from the analyzed dataset increases the reliability of recognition [1].

This research is funded by the Russian Foundation for Basic Research, grant 14-010-00039.

- [1] Borisova, I. A., Kutnenko O. A. 2015. Outliers detection in datasets with misclassified objects. *J. Mach. Learn. Data Anal.* 1(11):1632–1641. jmla.org/papers/doc/2015/no11/Borisova2015outliers.pdf.

Структурное обучение при порождении моделей

Варфоломеева Анна Андреевна varfaanna@gmail.com

Бочкарев Артем Максимович artem.bochkarev@phystech.edu

Москва, Московский физико-технический институт (ГУ)

Решается проблема порождения глобальной ранжирующей функции в задаче информационного поиска. Требуется найти такую ранжирующую функцию, которая определяла бы релевантность различных документов поступающим запросам. Для решения задачи поиска суперпозиции глобальной ранжирующей функции используется метод прогнозирования структуры функциональной зависимости. Предполагается, что функциональная зависимость существенно нелинейна и является суперпозицией элементарных функций, при этом делается ограничение на максимальную сложность модели. Дерево суперпозиции представляется в виде матрицы. В таком виде задача сводится к задаче структурного обучения — задаче восстановления структуры объекта по его описанию [1].

Для обучения метода структурного обучения получена выборка «подколлекция–модель» с помощью генетического алгоритма над множеством элементарных функций. Рассмотрен ряд методов прогнозирования структуры функции по матрице вероятности переходов, их работа протестирована на синтетических данных. Найденная глобальная ранжирующая функция протестирована при помощи функционала MAP на полной коллекции документов. Для проведения эксперимента использована выборка, состоящая из множеств коллекций документов конференции EURO 2010.

- [1] *Бочкарев А. М., Варфоломеева А. А.* Структурное обучение при порождении моделей // Машинное обучение и анализ данных, 2015 (в печати).

Structured learning for model generation

*Varfolomeeva Anna**

varfaanna@gmail.com

Bochkarev Artem

artem.bochkarev@phystech.edu

Moscow, Moscow Institute of Physics and Technology (State University)

The problem of the generation of the global ranking function in the task of information retrieval has been suggested. It is required to find a ranking function, which would determine the relevance of various documents incoming queries. To solve the problem of searching the superposition of the global ranking function, a method of forecasting the structure of functional dependencies has been used. It is assumed that the functional dependencies are nonlinear and ranking function is a superposition of elementary functions. The maximum complexity of the model is taken to be limited and a superposition tree is represented as a matrix. In this way, the problem reduces to the problem of structured learning — the problem of reconstructing the structure of the object by its description [1].

For training structured learning method, sample “subcollection – model” was received using a genetic algorithm on the set of elementary functions. A number of methods for predicting the structure function has been considered on the probability of transition, these methods are tested on synthetic data. Found global ranking function is tested using functional MAP on full collection of documents. To perform the experiment, a sample consisting of theses of documents of the conference EURO 2010 has been used.

- [1] Bochkarev, A., and Varfolomeeva A. 2015 (in press). Structure learning for model generation. *J. Mach. Lern. Data Anal.*

О метрических свойствах медианы Кемени

*Двоенко Сергей Данилович**

dsd@tsu.tula.ru

Пшеничный Денис Олегович

denispshenichny@yandex.com

Тула, Тульский государственный университет

Рассмотрена новая задача построения медианы Кемени с метрическими свойствами. Известно, что получение ранжирований означает выполнение измерений в так называемых менее мощных (качественных) шкалах. При согласовании экспертных мнений требуется получить общую ранжировку, наименее отличающуюся от остальных и имеющую смысл группового мнения. Медиана Кемени является эквивалентом среднего в шкалах (квази)порядков и свободна от ряда противоречий, связанных с выявлением групповых мнений по правилу большинства (парадокс Эрроу). Известный локально-оптимальный алгоритм построения медианы Кемени предполагает вычисление парных расстояний между ранжированиями и вычисление так называемой матрицы потерь.

Считая, что ранжирования, представленные парными расстояниями между собой, образуют множество, погруженное в евклидово метрическое пространство, получаем возможность определить средний элемент как центр такого множества. Очевидно, что такой центральный элемент также является ранжированием и должен иметь такой же смысл, как и медиана Кемени.

В общем случае найденный средний элемент отличается своими расстояниями от соответствующих расстояний медианы Кемени до остальных элементов множества. Разработана процедура формирования скорректированной матрицы потерь для построения метрической медианы Кемени, совпадающей со средним элементом данного множества. Такое ранжирование совпадает с классической медианой Кемени, подтверждая ее метричность, или отличается от нее, если метрические нарушения в конфигурации исходного множества ранжирований оказались значительными.

Работа поддержана грантом РФФИ № 15-07-02228.

- [1] Двоенко С. Д., Пшеничный Д. О. О метрических свойствах медианы Кемени // Машинное обучение и анализ данных, 2015. Т. 1. № 11. С. 1619–1631. jml.da.org/papers/doc/2015/JMLDA2015no11.pdf.

On metric characteristics of the Kemeny's median

*Dvoenko Sergey**

dsd@tsu.tula.ru

Pshenichny Denis

denispshenichny@yandex.com

Tula, Tula State University

The novel problem of the metric Kemeny's median is under investigation. It is well-known that the rankings mean the measurements in so-called less powerful (qualitative) scales.

For aggregating of experts' opinions, it is necessary to find the final ranking, which is the least different from others and represents the group opinion. The Kemeny's median appears to be the good idea of the average for scales of (quasi-)orderings and is free of some contradictions concerning the building of a group opinion based on the majority rules (Arrow's paradox). The well-known locally optimal algorithm to find the Kemeny's median depends on pairwise distances between rankings and calculates the so-called loss matrix.

It is assumed that the rankings represented by pairwise distances between them are immersed as a set in some Euclidean metric space. According to it, one can define the average element as the center of this set. It is evident that such the central element is also ranking and needs to be similar to the Kemeny's median.

In general, such an average element differs by its distances from the corresponding distances of the Kemeny's median to other set elements. The procedure is developed to build the modified loss matrix and to find the metric Kemeny's median, which coincides with the average element of the given set. Such ranking coincides with the classic Kemeny's median and proves its metric property, or differs from it, if the metric violations in the set configuration appear to be significant.

This research is funded by the Russian Foundation for Basic Research, grant 15-07-02228.

- [1] Dvoenko, S., and Pshenichny D. 2015. On metric characteristics of the Kemeny's median. *J. Mach. Learn. Data Anal.* 1(11):1619–1631. jmla.org/papers/doc/2015/JMLDA2015no11.pdf.

Применение байесовской регуляризации при обучении распознавания сигналов на основе метода опорных векторов

*Квасов Андрей Федорович**

kvandrew11@gmail.com

Красоткина Ольга Вячеславовна

krasotkina@tsu.tula.ru

Москва, Московский государственный университет
им. М. В. Ломоносова

Описаны теоретические и практические результаты, полученные при решении задач распознавания рукописных символов с помощью предложенного авторами критерия гладкости и метода опорных векторов (support vector machine — SVM).

Рассмотрены проблемы получения признакового пространства объектов для SVM, являющихся непрерывными сигналами с нефиксированным количеством измерений.

Производится поиск подходящих ограничений на априорное распределение обучающей выборки, количество объектов в которой сравнимо с размерами полученного признакового пространства.

Подробно освещен эвристический критерий гладкости, используемый как регуляризатор в SVM, его вероятностная интерпретация и результаты использования в задаче распознавания последовательностей измерений как в исходном пространстве генеральной совокупности, так и в метрическом пространстве объектов.

Также проведены исследования метода выравнивания сигналов Dynamic Time Warping, использованного для рассмотрения задачи распознавания рукописных символов в метрическом пространстве [1].

Представлены результаты работы классификатора на модельных и реальных данных, взятых из выборки UJI Pen Characters Data Set.

- [1] *Квасов А. Ф., Красоткина О. В.* Применение байесовской регуляризации при обучении сигналов на основе метода опорных векторов // Машинное обучение и анализ данных, 2015 (в печати).

Application of Bayesian regularization in signal recognition based on Support Vector Machine

*Kvasov Andrew**

`kvandrew11@gmail.com`

Krasotkina Olga

`krasotkina@tsu.tula.ru`

Moscow, Lomonosov Moscow State University

The paper describes theoretical and practical results obtained by solving the problem of online handwritten character recognition, using the smooth criterion proposed by the authors and included in Support Vector Machine (SVM).

The main objective was an investigation on objects' feature space used in SVM, which consists of continuous measurement sequences, i. e., signals, with nonfixed length.

The role of appropriate constraints, arising from the prior probability of the training sample which is characterized by the number of objects comparable to the size of the feature space, has been examined.

Detailed concept of the heuristic smoothness' criterion, used as a regularizer in SVM, with its probabilistic interpretation and use in signals' recognition, viewed both in the original space of the statistical population and in its metric space, is introduced.

Furthermore, this study includes an exploration of the signals' alignment method, called Dynamic Time Warping, used for construction of the signals' metric space [1].

The results obtained from the online handwriting character classification are performed on simulated data and real data taken from the samples of UJI Pen Characters Data Set.

- [1] Kvasov, A., and Krasotkina O. 2015 (in press). Application of Bayesian regularization in signal recognition based on Support Vector Machine. *J. Mach. Learn. Data Anal.*

Управление селективностью аномальных наблюдений в обучении одноклассовому распознаванию образов

Ларин Александр Олегович^{1*}

ekzebox@gmail.com

*Середин Олег Сергеевич*²

oseredin@yandex.ru

*Моттль Вадим Вячеславович*³

vmottl@ccas.ru

¹Москва, Московский физико-технический институт (ГУ)

²Тула, Тульский государственный университет

³Москва, Вычислительный центр им. А. А. Дородницына РАН

Наиболее популярным методом решения задачи обучения распознаванию образов по обучающей выборке, представленной объектами только лишь одного класса, является метод описания данных опорными векторами (SVDD – Support Vector Data Description), предложенный Д. Тэксом. Суть метода заключается в описании большинства векторов обучающей совокупности гиперсферой за исключением относительно небольшого числа далеких редких выбросов (outliers).

Основной недостаток данного метода в том, что он не обеспечивает корректной селекции нетипичных объектов обучающей совокупности. Это связано с тем, что исходная постановка задачи Тэкса не является геометрически верной, так как величина штрафа допустимого выхода объектов обучающей выборки за пределы описывающей гиперсферы несоизмерима с расстоянием до ее центра в оптимизационной задаче.

Цель данной работы — модификация метода SVDD с целью устранить его неадекватность природе многих прикладных задач [1]. Проблема заключается в том, что такая корректировка постановки задачи классификации приводит к необходимости оптимизации выпуклого, но недифференцируемого критерия. В данной работе предложен подход к решению оптимизационных задач такого типа.

Работа поддержана грантами РФФИ 14-07-00527-а и 14-37-50786-мол-нр.

- [1] *Larin A., Seredin O., Kopylov A., Kuo S.-Y., Huang S.-C., Chen B.-H.* Parametric representation of objects in color space using one-class classifiers // Machine learning and data mining in pattern recognition / Ed. P.Perner. — Lecture notes in computer science ser. — Switzerland: Springer International Publishing, 2014. Vol. 8556. P. 300–314.

Outlier selectivity control in one-class pattern recognition learning

Larin Alexander^{1*}

ekzebox@gmail.com

*Seredin Oleg*²

oseredin@yandex.ru

*Mottl Vadim*³

vmottl@ccas.ru

¹Moscow, Moscow Institute of Physics and Technology (State University)

²Tula, Tula State University

³Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

The Support Vector Data Description (SVDD) method proposed by D. Tax is one of the most popular methods to solve the problem of pattern recognition learning in training sets that contain objects of only one class. The essence of the method is description of the majority of objects in the training set by a hypersphere, except a relatively small amount of outliers. The main drawback of that method consists in its inability to provide correct selection of atypical objects. This owes to the fact that the original statement of the optimization problem proposed by D. Tax is geometrically incorrect, because the penalty for a point beyond the hypersphere is incommensurable with its distance to the hypersphere's center. The aim of this paper is to modify the SVDD method so as to eliminate its inadequacy to many applications [1]. The main problem is that such a modification leads to the necessity to minimize a convex but nondifferentiable training criterion. In this work, an approach to solve optimization problems of this kind is proposed.

This research is funded by the Russian Foundation for Basic Research, grants 14-07-00527-a and 14-37-50786-jung.

- [1] Larin, A., Seredin O., Kopylov A., Kuo S.-Y., Huang S.-C., and Chen B.-H. 2014. Parametric representation of objects in color space using one-class classifiers. *Machine learning and data mining in pattern recognition*. Ed. P. Perner. Lecture notes in computer science ser. Switzerland: Springer International Publishing. 8556:300–314.

Исследование и выбор регрессионных моделей с учетом погрешности как зависимых, так и независимых переменных

Рудой Георгий Игоревич

0xd34df00d@gmail.com

Москва, Московский физико-технический институт (ГУ)

В ходе анализа результатов физического эксперимента требуется восстановить функциональную зависимость между измеряемыми величинами, при этом необходима возможность экспертной интерпретируемости соответствующей зависимости.

Рассматривается класс существенно нелинейных параметрических суперпозиций элементарных функций, множество которых может быть либо заранее предложено экспертами, либо порождено, например, методом символьной регрессии. Однако в ходе физического эксперимента измеряемые величины как независимые, так и зависимые, известны лишь с некоторой конечной точностью. Соответственно требуется не только выбрать регрессионную модель, минимизирующую сумму квадратов регрессионных остатков, но и оценить зависимость вариации ее параметров от вариации входных данных в рамках некоторых предположений о погрешностях измерения.

Кроме того, непосредственно в процессе построения модели необходимо учитывать погрешности измерения независимых переменных, для чего стандартный функционал среднеквадратичной ошибки не подходит.

В настоящей работе предложен критерий устойчивости моделей, описывающий зависимость вариации параметров модели от вариации обучающих данных, и вводится функционал ошибки, учитывающий погрешности измерения независимых переменных, и соответствующий алгоритм оптимизации, основанный на алгоритме Левенберга–Марквардта. Предложенный метод также может быть применен в случае различия погрешности определения каждой из физических величин в различных точках [1].

[1] *Рудой Г. И.* О возможности применения методов Монте-Карло в анализе нелинейных регрессионных моделей // Сибирский журнал вычислительной математики, 2015 (в печати).

Model selection and analysis with respect to measurement errors in both free and dependent variables

*Georg Rudoy**

0xd34df00d@gmail.com

Moscow, Moscow Institute of Physics and Technology (State University)

A functional dependency between measured data is to be found during physical experiment analysis. It is desirable for the dependency to be interpretable by experts. For this, a set of nonlinear parametric superpositions of elementary functions is considered (either inductively generated via, for example, symbolic regression, or proposed by experts) as the set of candidate regression models.

However, during the physical experiment, the measured values both independent and dependent are known only to a finite precision. Thus, it is also required to estimate the dependency of variation of the selected model parameters on the variation of the learning data, accounting for measurement errors.

Moreover, during the model selection process (and, in particular, during parameters optimization), the possible measurement errors of independent variables should be taken into account. The standard mean-square error functional does not fit this task since it assumes zero measurement error of these variables.

This work proposes a model selection criterion called *model stability*, describing the dependency of model parameters variation on learning data variation, and a functional accounting for measurement errors in both dependent and independent variables along with the optimization method based on the Levenberg–Marquardt algorithm. The proposed criterion and method can be also used in case of different measurement errors in various data points [1].

[1] Rudoy, G. 2015 (in press). On applying Monte-Carlo methods to analysis of non-linear regression models. *Numerical Anal. Appl.*

Регрессионная модель, основанная на выпуклых комбинациях, максимально коррелирующих с откликом

Сенько Олег Валентинович

senkoov@mail.ru

*Докукин Александр Александрович**

dalex@ccas.ru

Москва, Федеральный исследовательский центр «Информатика и управление» РАН

Представлен новый регрессионный метод, основанный на построении оптимальных выпуклых комбинаций простых линейных регрессий по методу наименьших квадратов (МНК), построенных по исходным объясняющим переменным. Было показано, что на самом деле описанный регрессионный метод эквивалентен модификации МНК, включающей дополнительное требование о совпадении знака регрессионного параметра с коэффициентом корреляции между соответствующей объясняющей переменной и откликом. Описан метод построения оптимальных выпуклых комбинаций, основанный на концепции несократимых и нерасширяемых наборов. Представлены результаты экспериментов по сравнению разработанного метода с известным алгоритмом glmnet. Эксперименты подтвердили эффективность разработанного метода [1].

Работа поддержана грантом РФФИ № 14-07-00819.

- [1] *Сенько О. В., Докукин А. А.* Регрессионная модель, основанная на выпуклых комбинациях, максимально коррелирующих с откликом // Журнал вычислительной математики и математической физики, 2015. Т. 55. № 3. С. 530–544.

Regression model based on convex combinations best correlated with response

Senko Oleg

senkoov@mail.ru

*Dokukin Alexander**

dalex@ccas.ru

Moscow, Federal Research Center “Computer Science and Control”
of the Russian Academy of Sciences

A new regression method based on constructing optimal convex combinations of simple linear regressions of the least squares method (LSM) built from original regressors is presented. It is shown that in fact, this regression method is equivalent to a modification of the LSM including the additional requirement of the coincidence of the sign of the regression parameter with that of the correlation coefficient between the corresponding regressor and the response. A method for constructing optimal convex combinations based on the concept of nonexpandable irreducible ensembles is described. Results of experiments comparing the developed method with the known glmnet algorithm are presented, which confirm the efficiency of the former [1].

This research is funded by the Russian Foundation for Basic Research, grant 14-07-00819.

- [1] Senko, O., and Dokukin A. 2015. Regression model based on convex combinations best correlated with response. *Comput. Math. Math. Phys.* 55(3):526–539.

Метрическая гипотеза компактности и метод потенциальных функций в теории обучения

Середин Олег Сергеевич^{1*}

oseredin@yandex.ru

*Моттль Вадим Вячеславович*²

vmottl@ccas.ru

¹Тула, Тульский государственный университет

²Москва, Вычислительный центр им. А. А. Дородницына РАН

В качестве фундаментального принципа машинного обучения рассматривается гипотеза компактности, сформулированная Э. М. Браверманом в 1960 г. Она состоит в предположении наблюдателя, что объекты реального мира с близкими значениями скрытой целевой характеристики близки и по своим наблюдаемым свойствам. Мы понимаем гипотезу компактности в метрическом смысле, интерпретируя сходство в пространстве наблюдений как малое значение некоторой метрики, определенной на множестве объектов реального мира. Чтобы успешно решать задачи восстановления зависимости между наблюдаемой и скрытой характеристиками объекта, наблюдатель должен выбрать в пространстве наблюдений метрику, удовлетворяющую гипотезе компактности. Математический и алгоритмический аппарат обучения в произвольном метрическом пространстве основан на его мысленном погружении в псевдоевклидовое линейное пространство с последующим построением линейных, точнее, аффинных решающих правил, «почти» как в обычном евклидовом пространстве [1]. Однако принятая наблюдателем метрика не всегда обеспечивает приемлемое линейное псевдоевклидовое приближение объективно существующей зависимости. Предлагаемое параметрическое семейство преобразований исходной метрики, являющееся обобщением семейства потенциальных функций, разработанного Э. М. Браверманом для конечномерных евклидовых пространств, практически стирает различие между линейными и нелинейными решающими правилами.

Работа поддержана грантами РФФИ 14-07-00527-а и 14-07-00661-а.

- [1] Середин О. С., Абрамов В. И., Моттль В. В. Аффинные операции в псевдоевклидовом линейном пространстве // Известия ТулГУ, Естественные науки. — Тула: Изд-во ТулГУ, 2014. Вып. 3. С. 178–196.

The metric-based compactness hypothesis and method of potential functions for machine learning

Seredin Oleg^{1*}
*Mottl Vadim*²

oseredin@yandex.ru
vmottl@ccas.ru

¹Tula, Tula State University

²Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

The compactness hypothesis formulated by Emmanuel Braverman in 1960 is considered as the fundamental principle of machine learning. It consists in the observer's assumption that real-world objects having similar values of the hidden goal characteristics should be similar to each other also by their observable properties. We keep to the metric-based understanding of the compactness hypothesis, namely, consider the pairwise observation similarity as a small value of some metric defined on the set of real-world objects. It is up to the observer to find a metric that satisfies the compactness hypothesis. The mathematical and algorithmic framework of learning in an arbitrary metric space is based on its mental embedding into a pseudo-Euclidean linear space supplied with linear, to be more exact, affine decision rules, just "almost" as in a usual Euclidean linear space [1]. However, it happens often that the metric initially accepted by the observer does not provide good linear pseudo-Euclidean approximation to the actual dependence in the set of real-world objects. A parametric family of metric transformations, which is proposed in the present work as a generalization of Braverman's family of potential function onto metric spaces, virtually eliminates the difference between linear and nonlinear decision rules.

This research is funded by the Russian Foundation for Basic Research, grants 14-07-00527-a and 14-07-00661-a.

- [1] Seredin, O. S., Abramov V. I., and Mottl V. V. 2014. Affinnye operatsii v psevdoevklidovom lineynom prostranstve [Affine operations in a pseudo-Euclidean linear space]. *Proceedings of Tula State University, Natural Sciences*. Tula: Tula State University Publ. 3:178–196.

Отбор признаков в задаче классификации в потоках данных при смещении решающего правила

*Турков Павел Анатольевич**

pavel.turkov@gmail.com

Красоткина Ольга Вячеславовна

o.v.krasotkina@yandex.ru

Тула, Тульский государственный университет

В области распознавания образов существует множество задач классификации, связанных с потоками данных. В данной статье остановимся на таких важных особенностях данной проблемы, как бесконечная длина обучающей последовательности, смещение решающего правила и отбор признаков. Так как поток данных теоретически бесконечен, хранение и использование для обучения всех ранее полученных данных является невозможным. Смещение решающего правила является распространенным осложнением в задачах обучения в потоках данных, которое возникает в результате наличия изменений в наблюдаемых объектах. В литературе, посвященной восстановлению зависимостей в потоках данных, встречаются решения только для первых двух проблем, тогда как задача отбора признаков остается нерассмотренной. К сожалению, применение классических методов отбора признаков в процессе обучения осложнено смещением решающего правила. Представленная статья предлагает новый метод для отбора признаков, основанный на байесовском подходе к задаче распознавания образов в потоках данных. Предложена иерархическая вероятностная модель для оценки параметров решающего правила, позволяющая отбирать признаки наряду с оцениванием параметров решающего правила. Предложенный критерий является несмещенным, обладает хорошими обобщающими и прогнозирующими свойствами, его максимальный риск стремится к конечному значению. Эффективность предложенного подхода проиллюстрирована в ходе экспериментов на тестовых и реальных данных [1].

- [1] *Turkov P., Krasotkina O.* Feature selection in the data stream classification under concept drift // J. Mach. Learn. Data Anal., 2015 (в печати).

Feature selection in the data stream classification under concept drift

*Turkov Pavel**

pavel.turkov@gmail.com

Krasotkina Olga

o.v.krasotkina@yandex.ru

Tula, Tula State University

Data stream classification poses many challenges to the data mining community. In this paper, such major challenges, namely, infinite length, concept drift, and feature selection, are addressed. Since a data stream is theoretically infinite in length, it is impractical to store and use all the historical data for training. Concept drift is a common phenomenon in data streams, which occurs as a result of changes in the underlying concepts. Most existing data stream classification techniques address only the first two challenges and ignore the feature selection. The concept drift makes difficulties for applying classical feature selection methods in the learning procedure. This paper proposes the new Bayesian framework to feature selection in pattern recognition problem under concept drift. The hierarchical probability model is suggested for decision rule parameters estimation with sparse regularization. The proposed shrinkage criterion is unbiased, has grouping and oracle properties, and its maximal risk diverges to a finite value. Experimental results show that the proposed framework is competitive on both simulated data and publicly available real data [1].

- [1] Turkov, P., and Krasotkina O. 2015 (in press). Feature selection in the data stream classification under concept drift. *J. Mach. Learn. Data Anal.*

Новые расстояния и меры достоверности для формул логики Лукасевича в кластеризации логических высказываний базы знаний

Викентьев Александр Александрович^{1,2} vikent@math.nsc.ru

Фефелова Валерия Валерьевна^{2*} lexiyaren@yandex.ru

¹Новосибирск, Институт математики им. С. Л. Соболева СО РАН

²Новосибирск, Новосибирский государственный университет

Рассмотрена одна из актуальных задач, возникшая на стыке теории анализа данных, искусственного интеллекта и математической логики — задача анализа экспертных логических высказываний из базы знаний. Данное направление развивалось в работах Г. С. Лбова, Н. Г. Загоруйко, А. А. Викентьева и их коллег. В данной работе рассматриваются логические высказывания экспертов, представленные в виде формул n -значной логики Лукасевича. Понятно, что различные высказывания несут в себе разное количество информации. Следовательно, возникает вопрос об их сравнении по информативности и, как следствие, их ранжировании. Для кластеризации знаний, построения решающих функций на основе формул нужно ввести расстояния между формулами. С привлечением теории моделей определяются новые расстояния между формулами, обобщающие известные. Доказаны свойства метрики для расстояний и меры нетривиальности; они учитывают многозначность, схожи со свойствами величин предыдущих расстояний для логик Лукасевича, отвечают на вопросы Г. С. Лбова и применяются в алгоритмах кластеризации формул. Авторами введены новые расстояния и меры нетривиальности для формул n -значной логики Лукасевича, доказаны их свойства. Также был адаптирован иерархический алгоритм кластеризации [1].

Приведены примеры кластеризации для различных n и показаны их различия, написано рабочее java-приложение.

Работа выполнена при поддержке РФФИ, проекты 14-07-00851а и 14-07-00249а.

- [1] *Vikent'ev A. A.* Concerning distances and degrees of uncertainty for many-valued expert statements and application of those concepts in pattern recognition and clustering // *Pattern Recogn. Image Anal.*, 2014. Vol. 24. No. 4. P. 489–501.

New distances and invalidation measures for formulas of Lukasiewicz logic in clusterization of logical propositions of a database

Vikent'ev Alexandr^{1,2}

vikent@math.nsc.ru

Fefelova Valerija^{2*}

lexsiyaren@yandex.ru

¹Novosibirsk, Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences

²Novosibirsk, Novosibirsk State University

A topical problem on the border between data analysis, artificial intelligence, and mathematical logic has been considered: the analysis of databases of expert's statements. This area was developed by G. S. Lbov, N. G. Zagorujko, A. A. Vikent'ev, and their colleagues.

The present authors consider expert's logical statements represented by formulas of the n -valued formulas of Lukasiewicz logic. Clearly, different propositions can carry different amounts of information and it follows that the ranking of these statements by their informativeness becomes an interesting problem. The distances between formulas are considered as one of the tools for clusterization of knowledge and constructing deciding functions. Using model theory, new distances between formulas can be produced which generalize the already known ones. Some properties of metrics for such distances and of the measure of nontriviality have been proven; they take into account the many-valuedness of the logic, are similar to the previously studied distances for Lukasiewicz logic, answer some questions by G. S. Lbov, and are used in the clusterization algorithms for formulas. The authors introduce new distances and nontriviality measures for the formulas of the n -valued Lukasiewicz logic and prove some of their properties. Also, the hierarchic clusterization algorithm has been adapted [1].

The authors present some examples for various n and discuss the differences between them. Also, a working java application has been developed.

The research was supported by the Russian Foundation for Basic Research, projects 14-07-00851a and 14-07-00249a.

- [1] Vikent'ev, A. 2014. Concerning distances and degrees of uncertainty for many-valued expert statements and application of those concepts in pattern recognition and clustering. *Pattern Recogn. Image Anal.* 24(4):489–501.

Вопросы эффективности логических корректоров

Дюкова Елена Всеволодовна

edjukoval@mail.ru

Журавлёв Юрий Иванович

zhuravlev@ccas.ru

*Прокофьев Пётр Александрович**

p_prok@mail.ru

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

Одним из основных понятий, используемых для построения корректных распознающих процедур, является понятие элементарного классификатора (эл.кл.). Элементарный классификатор — это элементарная конъюнкция, определенная на целочисленных признаковых описаниях объектов. Если на описании некоторого объекта элементарная конъюнкция обращается в единицу, то говорят, что этот объект содержит данный эл.кл. Элементарный классификатор различает два объекта, если один объект содержит, а другой не содержит данный эл.кл. Элементарный классификатор называется корректным для класса K , если любые два обучающих объекта, один из которых принадлежит K , а другой не принадлежит K , одновременно не содержат данный эл.кл.

Классический логический анализ данных в распознавании основан на построении семейств из корректных эл.кл. Существуют сложные задачи, когда не удается найти достаточное количество информативных корректных эл.кл. Подобная ситуация возникает, например, в случае, когда каждый признак имеет слишком много значений. Один из способов решения проблемы — использование распознающих процедур, основанных на построении семейств корректных наборов эл.кл. (логических корректоров). Набор эл.кл. U называется корректным для класса K , если для любой пары обучающих объектов, в которой один объект из K , другой не из K , в U существует эл.кл., различающий эти объекты.

В [1] приводятся новые результаты, касающиеся повышения распознающей способности и скорости обучения логических корректоров. Строится модель логического корректора, базирующаяся на более общем понятии корректного набора эл.кл.

Работа частично поддержана грантами РФФИ № 13-01-00787-а, и № 14-07-00819-а и грантом президента РФ НШ-4908.2014.1.

[1] Дюкова Е. В., Журавлёв Ю. И., Прокофьев П. А. Методы повышения эффективности логических корректоров // Машинное обучение и анализ данных, 2015 (в печати).

The effectiveness of logical correctors

Djukova Elena

edjukoval@mail.ru

Zhuravlev Yury

zhuravlev@ccas.ru

*Prokofjev Petr**

p_prok@mail.ru

Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

One of the key concepts used to build the correct recognition procedures is the concept of elementary classifier (EC). Elementary classifier is an elementary conjunction defined on integer attributive descriptions of objects. If EC refers to 1 by the description of the object S , one say that object S contains EC. The objects S' and S'' are discriminated by EC, if exactly one of these objects contains it. Elementary classifier is *correct* for the class K , if between any training objects $S' \in K$ and $S'' \notin K$ either S' or S'' does not contain it.

Classical logical analysis of data by correct recognition is based upon the construction of correct EC families. There are challenges that cannot find a sufficient number of correct informative EC. Such situation arises, for example, in the task where attribute domains are too large. One way to solve the problem is to build the recognition procedures based on the construction of correct EC sets families (*logical correctors*). Set of EC U is *correct* for the class K if for any training objects $S' \in K$ and $S'' \notin K$, there exists EC that belongs to U by which S' and S'' are discriminated.

There are some new results concerning the improvement of recognition power and learning speed of logical correctors in [1]. The model of the logical corrector based on a more general concept of a correct set of EC is built.

This research is funded by the Russian Foundation for Basic Research, grants 13-01-00787-a and 14-07-00819-a and the President's of Russia grant SS-4908.2014.1.

- [1] Djukova, E. V., Zhuravlev Yu. I., and Prokofjev P. A. 2015 (in press). Methods to improve the effectiveness of logical correctors. *J. Mach. Learn. Data Anal.*

Об эффективных параллельных асимптотически оптимальных алгоритмах дуализации

*Дюкова Елена Всеволодовна*¹ edjukova@mail.ru
Никифоров Андрей Геннадьевич^{2*} ankifor@gmail.com

¹Москва, Вычислительный центр им. А. А. Дородницына РАН

²Москва, Московский государственный университет
им. М. В. Ломоносова, Факультет ВМК

Одной из фундаментальных задач дискретной математики является дуализация, которая в матричной формулировке звучит как задача построения (перечисления) всех неприводимых покрытий булевой матрицы.

Как правило, число решений дуализации растет экспоненциально с ростом размеров входных данных, поэтому актуальным является использование параллельных вычислений.

Основным результатом настоящей работы является развитие ранее предложенного авторами подхода к построению эффективных параллельных алгоритмов дуализации. Подход основан на статистических оценках объемов вычислительных подзадач.

В рамках рассматриваемого подхода разработана новая схема распараллеливания асимптотически оптимальных алгоритмов дуализации, позволяющая снизить временные затраты на получение статистических данных. Распределение вычислительных подзадач осуществляется согласно заранее составленному расписанию. Для составления указанного расписания устанавливается вид распределения случайной величины, используемой для оценки объемов подзадач, и оптимизируется уровень загрузки процессоров. Схема применена к распараллеливанию асимптотически оптимального алгоритма RUNC-M (Дюкова Е. В., Прокофьев П. А., 2014). Выявлены условия, при которых параллельная версия алгоритма RUNC-M демонстрирует достаточно равномерную загрузку процессоров и ускорение, близкое к максимальному [1].

Работа частично поддержана грантами РФФИ № 13-01-00787-а и № 14-07-00819-а и грантом президента РФ НШ-4908.2014.1.

[1] Дюкова Е. В., Никифоров А. Г. Об эффективном распараллеливании алгоритмов для дискретных перечислительных задач // Машинное обучение и анализ данных, 2015 (в печати). Т. 1. №. 13. jmla.org/papers/doc/2015/JMLDA2015no13.pdf.

On efficient parallel asymptotically optimal dualization algorithms

*Djukova Elena*¹

edjukova@mail.ru

Nikiforov Andrey^{2*}

ankifor@gmail.com

¹Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

²Moscow, Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics

Dualization is a fundamental problem in discrete mathematics and consists in enumeration of a given boolean matrix irreducible coverings.

As a rule, the number of irreducible coverings is exponential in the input matrix size. Therefore, it is essential to use parallel computing.

The basic result of the present paper is development of the approach to efficient parallel dualization algorithms elaboration, formerly suggested by the authors. The approach is based on statistical estimates for computational tasks size.

A new parallel scheme for asymptotically optimal dualization algorithms is developed, reducing time costs on statistical data collection. Task distribution is performed according to a timesheet set up in advance. For this purpose, a distribution of random variable, used for tasks size estimation, is fitted and the processor load level is optimized. A parallel scheme is applied to an asymptotically optimal algorithm RUNC-M (Djukova E., Prokofjev P., 2014). There are found conditions when parallel version of RUNC-M demonstrates a balanced processor load and almost maximal speedup [1].

This research is funded by the Russian Foundation for Basic Research, grants 13-01-00787-a and 14-07-00819-a and the President's of Russia grant SS-4908.2014.1.

- [1] Djukova, E., and Nikiforov A. 2015 (in press). On efficient parallelizing of the algorithms for discrete enumeration problems. *J. Mach. Learn. Data Anal.* 1(13). jmla.org/papers/doc/2015/JMLDA2015no13.pdf.

Об одной задаче кластеризации на графе

*Ильев Виктор Петрович**

iljev@mail.ru

Ильева Светлана Диадоровна

iljeva@mail.ru

Омск, Омский государственный университет

В задаче кластеризации требуется разбить заданное множество объектов на несколько подмножеств (*кластеров*) только на основе сходства объектов друг с другом. В теории распознавания образов и в машинном обучении задачи кластеризации относят к разделу *обучения без учителя*. Наряду с этим рассматриваются также *задачи кластеризации с частичным обучением*, в которых часть объектов изначально распределена по кластерам.

Одной из наиболее наглядных формализаций задач кластеризации взаимосвязанных объектов является *задача аппроксимации графа* (или *задача кластеризации на графе*), в которой требуется разбить множество вершин на кластеры с учетом реберной структуры. Граф называется *кластерным графом*, если каждая его компонента связности является полным графом. Обозначим через $\mathcal{M}_k(V)$ множество всех кластерных графов на множестве вершин V , содержащих ровно k непустых компонент связности, $2 \leq k \leq |V|$. Если $G_1 = (V, E_1)$ и $G_2 = (V, E_2)$ — графы на одном и том же множестве вершин V , то *расстояние* между ними определяется как $\rho(G_1, G_2) = |E_1 \setminus E_2| + |E_2 \setminus E_1|$.

Рассмотрим следующую задачу кластеризации на графе.

Задача \mathbf{A}_k^+ . Дан обыкновенный граф $G = (V, E)$, целое число k , $2 \leq k \leq |V|$, и множество $X = \{x_1, \dots, x_k\} \subseteq V$, где $x_i \neq x_j$ при $i \neq j$. Требуется найти такой граф $M^* \in \mathcal{M}_k(V)$, что

$$\rho(G, M^*) = \min_{M \in \mathcal{M}_k(V)} \rho(G, M),$$

где минимум берется по всем кластерным графам M , в которых $x_i \in V_i$. Здесь V_i — i -й кластер (компонента связности) графа M , $i = 1, \dots, k$.

Доказано, что задача \mathbf{A}_k^+ является NP-трудной для любого фиксированного $k \geq 2$. Для задачи \mathbf{A}_2^+ предложен полиномиальный 3-приближенный алгоритм [1].

Работа поддержана грантом РФФИ № 15-11-10009.

- [1] Ильев В. П., Ильева С. Д. О задачах кластеризации на графах // Вестник Омского университета. — Омск: Изд-во ОмГУ, 2015 (в печати).

On a graph clustering problem

*Il'ev Victor**

iljev@mail.ru

Il'eva Svetlana

iljeva@mail.ru

Omsk, Omsk State University

The objective of the *clustering problem* is to partition the objects (data elements) into a family of subsets (i.e., *clusters*) such that the objects within a cluster are more similar to one another than the objects in different clusters. In pattern recognition and machine learning, clustering methods fall under the section of *unsupervised learning*. At the same time, *semisupervised* clustering methods use limited supervision. For example, relatively few objects are labeled (i.e., are assigned to clusters), whereas a large number of objects are unlabeled.

One of the most visual formalizations of clustering is the *graph approximation problem* (or *graph clustering*), that is, grouping the vertices of a graph into clusters taking into consideration the edge structure of the graph. A graph is called a *cluster graph* if each of its connected components is a complete graph. Denote by $\mathcal{M}_k(V)$ the set of all cluster graphs on vertex set V consisting of exactly k nonempty connected components, $2 \leq k \leq |V|$. If $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ are the graphs on the same vertex set V , then the *distance* between them is defined as $\rho(G_1, G_2) = |E_1 \setminus E_2| + |E_2 \setminus E_1|$.

Let consider the following graph clustering problem.

Problem \mathbf{A}_k^+ . Given a graph $G = (V, E)$, an integer k , $2 \leq k \leq |V|$, and a set $X = \{x_1, \dots, x_k\} \subseteq V$ ($x_i \neq x_j$ unless $i = j$), find a graph $M^* \in \mathcal{M}_k(V)$ such that

$$\rho(G, M^*) = \min_{M \in \mathcal{M}_k(V)} \rho(G, M)$$

where minimum is taken over all cluster graphs M such that $x_i \in V_i$, $i = 1, \dots, k$, with V_i being the i th cluster (connected component) of graph M .

It has been proved that problem \mathbf{A}_k^+ is NP-hard for each fixed integer $k \geq 2$. For problem \mathbf{A}_2^+ , 3-approximation polynomial-time algorithm has been proposed [1].

This research is funded by the Russian Science Foundation, grant 15-11-10009.

- [1] Il'ev, V. P., and Il'eva S. D. 2015 (in press). O zadachakh klasterizatsii na grafakh [On graph clustering problems]. *Vestnik Omskogo Universiteta* [Bulletin of Omsk University].

Комбинаторные оценки переобучения одномерных пороговых классификаторов

*Ишкина Шаура Хабировна**

shaurushka@gmail.com

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

Одномерные пороговые классификаторы применяются для бинаризации данных при построении логических закономерностей и решающих деревьев, а также в качестве решающего правила в моделях классификации с вещественной дискриминантной функцией. В данной работе получены точные оценки обобщающей способности для данного семейства классификаторов в рамках комбинаторной теории переобучения.

Непосредственное вычисление комбинаторных оценок обобщающей способности по определению имеет экспоненциальную сложность. Предлагается алгоритм вычисления вероятности переобучения и ожидаемой частоты ошибок скользящего контроля за полиномиальное время $O(\ell^6)$, где ℓ — объем выборки. Рассматриваются два варианта алгоритма: для количественного признака, принимающего попарно различные значения на объектах выборки, и для номинального признака, индуцирующего вариационный ряд со связками.

Алгоритм основан на рекуррентном подсчете числа допустимых траекторий при блуждании по трехмерной сетке между двумя заданными точками с ограничениями специального вида.

Полученные оценки применены в качестве критерия отбора признаков для решения прикладной задачи диагностики заболеваний внутренних органов человека методами информационного анализа электрокардиосигналов по В. М. Успенскому. Показано, что при малых объемах обучающей выборки для некоторых заболеваний данные оценки позволяют получить менее сложную модель при более высокой точности диагностики [1].

Работа поддержана РФФИ, грант № 14-07-00847.

- [1] *Ишкина Ш. Х.* Комбинаторные оценки переобучения пороговых решающих правил // Уфимский математический журнал, 2015 (в печати).

Combinatorial bounds of overfitting for threshold decision rules

*Ishkina Shaura**

shaurushka@gmail.com

Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

One-dimensional threshold decision rules are widely used in supervised learning as data binarization functions, building blocks for decision trees and conjunction rules, and decision functions for real-valued classifiers. In this work, combinatorial generalization bounds have been obtained for learning threshold decision rules via empirical risk minimization.

The direct computation of the probability of overfitting in combinatorial theory of overfitting has an exponential time complexity in the size of the dataset ℓ . The author proposes an algorithm that computes an exact upper bound of the probability of overfitting in polynomial time $O(\ell^6)$. Two versions of the algorithm, both for continuous and discrete features have been considered.

The algorithms exploit the isomorphism between the set of partitions of the dataset into training and testing subsets that suffer from overfitting, and the set of paths of the constrained random walk on three-dimensional lattice. The algorithm calculates recursively the number of admissible paths that connect two fixed points of the lattice.

The overfitting bounds have been used as a feature selection criterion for Naïve Bayes linear classifier. The experiments were carried out on the set of 600 cardiosycle electrocardiogram records labeled by diagnoses of 18 internal diseases. The features were extracted as frequencies of 216 patterns of heart rate variability covering 4 neighboring cardiocycles, according to the information analysis of ECG-signals proposed by V. M. Uspenskiy. It is shown that combinatorial bounds lead to a more accurate and less complicated classification models [1].

This research is funded by the Russian Foundation for Basic Research, grant 14-07-00847.

[1] Ishkina Sh. 2015 (in press). Combinatorial bounds of overfitting for threshold decision rules. *Ufa Math. J.*

Задача отделимости (separation problem) опорных неравенств для задачи аппроксимации графа

Кононов Александр Вениаминович¹ alvenko@math.nsc.ru
 Симанчев Руслан Юрьевич^{2,3*} osiman@rambler.ru
 Уразова Инна Владимировна³ urazovainn@mail.ru

¹Новосибирск, Институт математики им. С. Л. Соболева СО РАН

²Омск, Омский научный центр СО РАН

³Омск, Омский государственный университет им. Ф. М. Достоевского

Пусть $\mathbf{K}_n = (V, E)$ — полный неориентированный граф без петель и кратных ребер. Остовный подграф $H \subset \mathbf{K}_n$ называется M -графом, если каждая его компонента связности является кликой. При этом полагаем, что одновершинный граф также клика. Множество всех M -графов в \mathbf{K}_n обозначим через $\mu(V)$.

Вектором инцидентий произвольного графа $D \subseteq K_n$ без изолированных вершин называется вектор $x^D \in R^E$ с компонентами

$$x_e^D = \begin{cases} 1 & \text{при } e \in ED; \\ 0 & \text{при } e \notin ED. \end{cases}$$

Многогранником M -графов будем называть множество

$$P_n = \text{conv} \{x^H \in R^E \mid H \in \mu(V)\}.$$

В [1] построен класс опорных неравенств (k -парашюты) относительно многогранника P_n , описаны условия, при которых полученные неравенства являются фасетными.

При использовании опорных неравенств в качестве отсекающих плоскостей на передний план выходит задача идентификации (separation problem), которая состоит в том, чтобы найти в заданном классе неравенство, отсекающее данную нецелочисленную точку, либо доказать, что в этом классе такого неравенства нет. В настоящей работе получен следующий результат.

Теорема. Задача идентификации неравенств, индуцированных k -парашютами, относительно многогранника задачи аппроксимации графов NP-трудна.

- [1] Симанчев Р. Ю., Уразова И. В. О гранях многогранника задачи аппроксимации графов // Дискрет. анализ и исследование операций, 2015. Т. 22. № 2. С. 86–101. http://math.nsc.ru/publishing/DAOR/2_15.html.

On separation problem of same support inequalities class for M -graphs polytope

*Kononov Alexandr*¹

alvenko@math.nsc.ru

Simanchev Ruslan^{2,3*}

osiman@rambler.ru

*Urazova Inna*³

urazovainn@mail.ru

¹Novosibirsk, Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences

²Omsk, Omsk Scientific Center of the Siberian Branch of the Russian Academy of Sciences

³Omsk, Omsk State University

Let $K_n = (V, E)$ is usual complete graph without loops and multiple edges with the vertex set V and edge set E . Spanning subgraph $H \subset K_n$ is called M -graph if each its connected component is a clique. It is assumed that one-vertex graph is clique too. The set of all M -graphs of K_n let denote $\mu(V)$.

Let R^E is the space of column vectors and axes of this space are in one-to-one correspondence with an edge E . For each subgraph $H \subseteq K_n$, let define its incidence vector as $(0, 1)$ -vector $x^H \in R^E$ with the components

$$x_e^H = \begin{cases} 1, & \text{if } e \in EH; \\ 0 & \text{otherwise.} \end{cases}$$

The M -graphs polytope is the set defined as

$$P_n = \text{conv} \{x^H \in R^E | H \in \mu(V)\}.$$

In [1], a class of support inequalities (k -parachutes) was described with respect to P_n . Moreover, the conditions under which these inequalities are facet have been also described.

The support inequalities have been used in the cutting plane procedures. Here, the separation problem becomes very important. The separation problem is as follows. Given a point in R^E and a class of support inequalities for P_n . Either find the inequality of this class which separates point and P_n or prove that there is no such inequality in this class. In this paper, the following result is obtained.

Theorem. The separation problem for k -parachute inequalities and M -graphs polytope is NP-hard.

- [1] Simanchev, R.Yu., and Urazova I.V. 2015. On the polytope faces of the graph approximation problem. *J. Appl. Ind. Math.* 9(2):283–291.

Дифференцирование классов при оценке не полностью описанного объекта

Юдин Валерий Н.^{1*}

yudin@ispras.ru

*Карпов Леонид Е.*¹

mak@ispras.ru

*Карпов Юрий Л.*²

y.l.karpov@yandex.ru

*Сметанин Юрий Г.*³

ysmetanin@rambler.ru

¹Москва, Институт системного программирования РАН

²Москва, Люксифт Россия

³Москва, Вычислительный центр им. А. А. Дородницына РАН

Предложен основанный на алгоритме вычисления оценок метод распознавания образов в условиях нефиксированного набора признаков, неполноты наборов признаков для отдельных объектов, разных наборов признаков для разных классов, возможного появления новых классов. Для оценки объектов используются проекции классов на подпространства признаков, значения которых известны для этих объектов. Классы, в пересечение которых попадает объект, образуют так называемый дифференциальный ряд объекта. Предложен подход к построению опорных множеств для отдельных классов на основе анализа прецедентов, определения меры близости для полностью и не полностью определенных векторов признаков и расширения обучающих выборок за счет добавления к ним случаев решения задачи распознавания. При построении опорных множеств в условиях неполноты описаний используется оригинальная процедура проецирования и выделения информативных подпространств признаков. Для реализации предложенного метода разработана система распознавания, способная к адаптации к решаемым классам задач: появлению новых классов и расширению спектра возможных значений признаков. В качестве примера рассмотрена задача дифференциальной диагностики в медицине в условиях ограниченных ресурсов и времени [1].

Работа поддержана грантами РФФИ № 15-01-02362 и № 15-07-02355.

- [1] *Юдин В. Н., Карпов Л. Е., Карпов Ю. Л., Сметанин Ю. Г.* Дифференцирование классов при оценке не полностью описанного объекта // Машинное обучение и анализ данных, 2015 (в печати).

Differential set of classes for estimation of not fully described objects

Yudin Valeriy^{1*}

yudin@ispras.ru

*Karpov Leonid*¹

mak@ispras.ru

*Karpov Yury*²

y.l.karpov@yandex.ru

*Smetanin Yury*³

ysmetanin@rambler.ru

¹Moscow, Institute for System Programming of the Russian Academy of Sciences

²Moscow, Luxoft Russia

³Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

A method of pattern recognition is suggested based on calculating estimations in case the set of object features is not fixed, the set of feature values for individual objects is not necessary complete, different classes are characterized by different sets of features, and new classes may appear during the process of data analysis. The aim is to recognize objects described by incomplete sets of features, that is, the projections of classes on subspaces of features whose value is known. The classes in the intersection of which the object is located form a so-called differential set of the object. Support sets are constructed for each class as a result of the analysis of its training examples (precedents). A measure of proximity is proposed suitable for both complete and incomplete feature vectors. This measure is determined using an original procedure of finding the projections of feature vectors on the subspace spanned on the features whose value is known. The training set can be expanded by adding the recognized samples. A pattern recognition system is developed for the implementation of the proposed method; the system is capable to adapt to the applied problems, e. g., it can add training samples and new classes and can expand the range of possible feature values. As an example of application, the problem of decision making in differential diagnosis in medicine is considered under limited resources and time [1].

This research was partly supported by the Russian Foundation for Basic Research, grants 15-01-02362 and 15-07-02355.

- [1] Yudin, V., Karpov L., Karpov Yu., and Smetanin Yu. 2015 (in press). Differential set of classes for estimation of not fully described objects. *J. Mach. Learn. Data Anal.*

2-симплекс призма — когнитивное средство принятия и обоснования решений в интеллектуальных динамических системах

Янковская Анна Ефимовна^{1,2,3,4,*}

ayyankov@gmail.com

*Ямшанов Артем Вячеславович*⁴

yav@keva.tusur.ru

*Кривдюк Наталья Михайловна*⁴

skratnat@gmail.com

¹Томск, ТГАСУ

²Томск, ТГУ

³Томск, СибГМУ

⁴Томск, ТУСУР

Идея применения n -симплексов и теорема для принятия и обоснования решений в интеллектуальных медицинских системах была предложена А.Е. Янковской в 1990 г. Впервые предлагается применять 2-симплекс призму (рис. 1), в сечениях которой расположены 2-симплексы, для принятия и обоснования решений в интеллектуальных динамических системах (ИДС), основанных как на тестовых методах распознавания образов, так и на методах нечеткой и пороговой логики, в ряде проблемных областей: медицина, образование, дорожное строительство и др. [1].

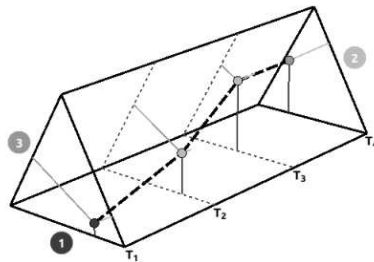


Рис. 1. 2-симплекс призма в ИДС

Работа поддержана грантами РФФИ (проекты №№ 13-07-00373, 13-07-98037, 14-07-00673) и частично РГНФ (проект № 13-06-00709).

- [1] Янковская А. Е., Ямшанов А. В., Кривдюк Н. М. 2-симплекс призма — когнитивное средство принятия и обоснования решений в интеллектуальных динамических системах // Машинное обучение и анализ данных, 2015 (в печати). Т. 1. № 13. jmla.org/papers/doc/2015/JMLDA2015no13.pdf.

2-simplex prism — a cognitive tool for decision-making and its justifications in intelligent dynamic systems

Yankovskaya Anna^{1,2,3,4}*

ayyankov@gmail.com

Yamshanov Artem⁴

yav@keva.tusur.ru

Krivdyuk Natalia⁴

skratnat@gmail.com

¹Tomsk, TSUAB

²Tomsk, TSU

³Tomsk, SSMU

⁴Tomsk, TUSUR

The idea of n -simplex application and the theorem for decision-making and its justifications for intelligent medical systems were proposed by A. Yankovskaya in 1990. Here, 2-simplex prism which contains 2-simplex disposed in geometrical sections for the first time is proposed (Fig. 1) for application at decision-making and its justifications in intelligent dynamic systems (IDS) based on test methods of pattern recognition and methods of fuzzy and threshold logics for different problem areas: medicine, education, road building, etc. [1].

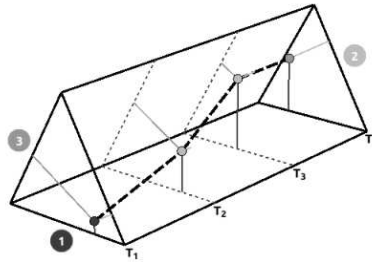


Fig. 1. 2-simplex prism in IDS

The work is supported by the Russian Foundation for Basic Research (projects 13-07-00373a, 13-07-98037-r_sibir_a, and 14-07-00673) and partially supported by the Russian Humanitarian Scientific Foundation (project 13-06-00709).

- [1] Yankovskaya, A., Yamshanov A., and Krivdyuk N. 2015 (in press). 2-simplex prism — a cognitive tool for decision-making and its justifications in intelligent dynamic systems. *J. Mach. Learn. Data Anal.* 1(13). jmla.org/papers/doc/2015/JMLDA2015no13.pdf.

Эффективные численные методы для дважды разреженных задач оптимизации большой размерности

Аникин Антон Сергеевич¹ anton.anikin@htower.ru
 Двуреченский Павел Евгеньевич² pavel.dvurechensky@gmail.com
 Гасников Александр Владимирович^{2,3,4*} gasnikov@yandex.ru
 Горнов Александр Юрьевич¹ gornov@icc.ru
 Камзолов Дмитрий Игоревич^{2,3} kamzolov-d@mail.ru
 Максимов Юрий Владимирович^{2,3,4} yurymaximov@iitp.ru

¹Иркутск, Институт динамики систем и теории управления СО РАН

²Москва, Институт проблем передачи информации РАН

³Москва, Московский физико-технический институт (ГУ),

Лаборатория структурных методов анализа данных

в предсказательном моделировании (ПреМоЛаб)

⁴Москва, Национальный исследовательский университет «Высшая школа экономики»

Рассматриваются задачи оптимизации большой размерности, записанные в виде:

$$\|Ax - b\|_2^2 + \lambda F(x) \rightarrow \min,$$

где $F(x)$ — простой сильно выпуклый регуляризатор (такой как $\|x\|_2^2$, $\|x\|_1$ или энтропия) и λ — числовой параметр.

Проблемы такого рода возникают часто во многих инженерных приложениях, равно как и в области машинного обучения, таких как ранжирование и задачи оптимизации в сложных сетях.

Рассматриваются только разреженные задачи. В зависимости от типа разреженности (по строкам / по столбцам, равномерная разреженность / разреженность в среднем) сравниваются различные современные алгоритмы и предлагаются некоторые новые идеи, позволяющие снизить их вычислительную сложность [1].

Работа поддержана грантом РФФИ 15-31-20571 мол_a_вед и Лабораторией структурных методов анализа данных в предсказательном моделировании факультета управления и прикладной математики МФТИ, грант правительства РФ 11.G34.31.0073.

[1] Гасников А. В., Двуреченский П. Е., Нестеров Ю. Е. Стохастические градиентные методы с неточным оракулом // Автоматика и телемеханика, 2015. arXiv:1411.4218, Nov. 2014. P. 1–53. <http://arxiv.org/pdf/1411.4218.pdf>.

Efficient numerical optimization methods for huge-scale sparse problems

*Anton Anikin*¹

anton.anikin@htower.ru

*Dvurechensky Pavel*²

pavel.dvurechensky@gmail.com

Gasnikov Alexander^{2,3,4,*}

gasnikov@yandex.ru

*Gornov Alexander*¹

gornov@icc.ru

Kamzolov Dmitry^{2,3}

kamzolov-d@mail.ru

Maximov Yury^{2,3,4}

yurymaximov@iitp.ru

¹Irkutsk, Institute for System Dynamics and Control Theory
of the Siberian Branch of the Russian Academy of Sciences

²Moscow, Institute for Information Transmission Problems
of the Russian Academy of Sciences

³Moscow, Moscow Institute of Physics and Technology (State University),
Laboratory of Structural Methods in Predictive Modeling (PreMoLab)

⁴Moscow, National Research University “Higher School of Economics”

The authors consider a huge-scale optimization problems in the form

$$\|Ax - b\|_2^2 + \lambda F(x) \rightarrow \min$$

where $F(x)$ is the composite convex function and λ is the small numerical parameter.

The problem typically arises in many engineering and machine learning problems, such as ranking or network topology optimization.

The matrix is supposed to be sparse. Depending on various sparsity conditions (column/row sparsity, uniform/average sparsity), the authors compare state-of-the-art algorithms and propose several novel ideas that allow to improve computational efficiency of numerical optimization methods [1].

The work is supported by the Russian Foundation for Basic Research, grant 15-31-20571 mol_a_ved, and by the Laboratory of Structural Methods in Predictive Modeling of the Department of Control and Applied Mathematics, Moscow Institute of Physics and Technology, Russian Government, grant 11.G34.31.0073.

[1] Gasnikov, A., Dvurechensky P, and Nesterov Yu. 2014. Stochastic gradient methods with inexact oracle. arXiv:1411.4218. P. 1–53.

Динамическая модель организации грузоперевозок

*Бекларян Лева Андреевич**

beklar@cemi.rssi.ru

Хачатрян Нерсес Карленович

nerses@cemi.rssi.ru

Москва, ЦЭМИ РАН

Исследуется модель, описывающая процесс грузоперевозок, реализуемый в рамках ряда технологий. Рассматриваются четыре варианта модели. Первый вариант модели описывает транснациональные транспортные грузоперевозки без выделенных начальной станции отправления и конечной станции распределения грузов. Этот вариант модели описывает грузоперевозки, для которых как первая, так и последняя станции не являются узловыми. Для таких грузоперевозок важно описать закон взаимодействия промежуточных станций. Второй вариант модели описывает транспортные грузоперевозки с выделенной начальной станцией отправления грузов. Этот вариант модели описывает грузоперевозки на протяженном участке пути, где начальная станция является узловой. Третий вариант модели описывает транспортные грузоперевозки с выделенными начальной станцией отправления и конечной станцией распределения грузов. Этот вариант модели описывает грузоперевозки на протяженном участке пути между двумя узловыми станциями. Четвертый вариант модели описывает транспортные грузоперевозки по круговой цепочке станций.

Для всех вариантов модели изучаются режимы грузоперевозок, удовлетворяющие заданной системе контроля. Такие режимы описываются решениями типа бегущей волны для нелинейного конечно-разностного аналога уравнения параболического типа. Описаны возможные режимы грузоперевозок, исследован вопрос устойчивости стационарных режимов [1].

Работа поддержана грантом РФФИ № 15-51-05011.

- [1] *Бекларян Л. А., Хачатрян Н. К.* Динамическая модель организации грузоперевозок // Машинное обучение и анализ данных, 2015 (в печати). Т. 1. № 14. jmlda.org/papers/doc/2015/JMLDA2015no14.pdf.

Dynamic model of the organization of cargo transportation

*Beklaryan Lieva**

beklar@cemi.rssi.ru

Khachatryan Nerses

nerses@cemi.rssi.ru

Moscow, CEMI RAS

The model describing the process of cargo transportation, realized through a number of technologies, is investigated. Four versions of the model are considered. The first version of the model describes the transnational cargo transportation without dedicated initial departure station and the final station cargo distribution. This version of the model describes the cargo, for which both the first and the last stations are not the nodes. For such cargo transportation, it is important to describe the rule of interaction of intermediate stations. The second version of the model describes the transport cargo with a dedicated initial departure station. This version of the model describes the cargo on the long section of the route where the initial departure station is nodular. The third version of the model describes the cargo transportation between dedicated initial departure station and final station. This version of the model describes the cargo on the long section of the route between the two node stations. The fourth version of the model describes the cargo transportation in a circular chain of stations.

For all versions of the model, the modes of freight satisfying given control system are studied. Such regimes are described by traveling wave type solutions of nonlinear finite-difference analogue of a parabolic equation. The possible modes of freight are described and the issue of stability of stationary regimes are investigated [1].

This research is funded by the Russian Foundation for Basic Research, grant 15-51-05011.

- [1] Beklaryan, L., and Khachatryan N. 2015 (in press). Dynamic model of the organization of cargo transportation. *J. Mach. Learn. Data Anal.* 1(14). jmla.org/papers/doc/2015/JMLDA2015no14.pdf.

О некоторых трудных задачах дискретной оптимизации, связанных с кластеризацией, покрытием и маршрутизацией

Гимади Эдуард Хайрутдинович^{1,2*} gimadi@math.nsc.ru

¹Новосибирск, Новосибирский государственный университет

²Новосибирск, Институт математики им. С. Л. Соболева СО РАН

Представлена тематика построения алгоритмов с оценками для решения некоторых труднорешаемых задач дискретной оптимизации, связанных с проблемами кластеризации, покрытия графов и маршрутизации.

Большая часть проблем состоит в поиске в графе нескольких дискретных реберно- или вершинно-непересекающихся структур (подграфов) с экстремальным суммарным весом в заданном полном реберно-взвешенном графе [1]. Особое внимание уделено результатам построения асимптотически точных алгоритмов.

Рассмотрены следующие задачи:

- (1) задача кластеризации на числовой оси;
- (2) задача кластеризации на сетевых моделях;
- (3) метрическая и квадратичная евклидова задача поиска m клик с максимальным суммарным весом;
- (4) поиск m векторов с максимальной нормой суммы в евклидовом пространстве: рандомизированный алгоритм;
- (5) TSP (Traveling Salesman Problem) подход к приближенному решению задачи покрытия графа m вершинно-несмежными циклами (цепями);
- (6) Евклидова задача MAX m -CYCLES COVER;
- (7) задача MIN m -CYCLES COVER на случайных входах из класса UNI $(0, 1)$;
- (8) задача поиска одного и двух остовных деревьев минимального веса при ограничении снизу на диаметр дерева;
- (9) m -Peripatetic Salesman Problem: поиск нескольких реберно-непересекающихся гамильтоновых циклов в полном графе.

Работа поддержана грантами РФФИ № 15-01-00976 и № 13-07-00070.

- [1] *Gimadi E.* Efficient algorithms with performance guarantees for some problems of finding several discrete disjoint subgraphs in complete weighted graph // *Appl. Math. Comput.*, 2015. Vol. 255. P. 84–91. doi:10.1016/j.amc.2014.11.037.

On some hard discrete optimization problems associated with clustering, covering, and rooting

Gimadi Edward^{1,2*}

`gimadi@math.nsc.ru`

¹Novosibirsk, Novosibirsk State University

²Novosibirsk, Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences

Several observations on efficient approximation algorithms with proven guarantees have been made for some discrete optimization problems that mainly associated with clustering, covering, and rooting. The problems considered are, in general, case NP-hard.

Most of the problems considered are to find multiple discrete disjoint structures (subgraphs) of extremal total weight in a given complete edge-weighted graph [1]. Particular attention has been paid to the results of designing asymptotically optimal algorithms, when a performance ratio tends to 1 with increasing size of the input data.

The following problems have been considered:

- (1) Clustering Problem on the real axis;
- (2) Clustering Problem in network models;
- (3) Metric and Quadratic Euclidean MAX m -Weighted Clique Problems: 2-approximation algorithm;
- (4) Euclidean MAX Vector Subset Problem: a randomized algorithm;
- (5) TSP (Traveling Salesman Problem) approach to construct an approximation algorithm for solving the problem m -CYCLES (CHAINS) COVER: covering a complete graph by m vertex-disjoint cycles (chains) of extremal total weight of edges;
- (6) Euclidean MAX m -CYCLES COVER;
- (7) Random MIN m -CYCLES COVER on instances UNI $(0, 1)$;
- (8) Minimum Spanning Tree Problem with diameter, bounded from below: cases of finding one and two disjoint spanning trees; and
- (9) m -Peripatetic Salesman Problem: finding several edge-disjoint Hamiltonian circuits in complete weighted graph.

This research is funded by the Russian Foundation for Basic Research, grants 15-01-00976 and 13-07-00070.

- [1] Gimadi, E. 2015. Efficient algorithms with performance guarantees for some problems of finding several discrete disjoint subgraphs in complete weighted graph. *Appl. Math. Comput.* 255:84–91. doi:10.1016/j.amc.2014.11.037.

Об асимптотически точных приближенных алгоритмах для некоторых трудных задач маршрутизации

Гимади Эдуард Хайрутдинович^{1,2} gimadi@math.nsc.ru

Цидулко Оксана Юрьевна^{1*} tsidulko.ox@gmail.com

¹Новосибирск, Институт математики им. С. Л. Соболева СО РАН

²Новосибирск, Новосибирский государственный университет

Рассмотрены три задачи, являющиеся модификациями классической задачи коммивояжера: задача об m коммивояжерах с различными весами маршрутов, задача покрытия графа m циклами и задача покрытия графа m цепями. Содержательно, первая задача состоит в том, чтобы в заданном полном (не)ориентированном графе $G = (V, E)$ с весовыми функциями $w_i : E \rightarrow \mathbf{R}_+$, $i = 1, \dots, m$, найти m реберно-непересекающихся гамильтоновых циклов минимального суммарного веса. Вторая и третья задачи заключаются в том, чтобы в полном взвешенном графе найти m циклов (цепей) минимального общего веса, таких что любая вершина графа G принадлежит ровно одному циклу (цепи).

Все три задачи NP-трудны в сильном смысле, поскольку к ним сводится классическая задача коммивояжера.

Предложены приближенные полиномиальные алгоритмы решения этих задач. Рассмотрено поведение этих алгоритмов в предположении, что веса ребер графа являются независимыми одинаково распределенными случайными величинами [1].

Найдены условия асимптотической точности алгоритмов в случаях, когда веса ребер графа G имеют равномерное распределение на отрезке $[a_n, b_n]$, $0 < a_n < b_n$, и усеченное показательное распределение с параметром β_n на интервале $[a_n, \infty)$, $a_n > 0$.

Работа выполнена при поддержке РФФИ, проекты 15-01-00976 и 13-07-00070.

- [1] *Гимади Э. Х., Истомин А. М., Рыков И. А., Цидулко О. Ю.* Вероятностный анализ приближенного алгоритма для решения задачи о нескольких коммивояжерах на случайных входных данных, неограниченных сверху // Тр. ИММ УрО РАН, 2014. Т. 20-2. С. 88–98. <http://mi.mathnet.ru/timm1061>.

On polynomial asymptotically optimal algorithms for some hard routing problems

Gimadi Edward^{1,2}

`gimadi@math.nsc.ru`

Tsidulko Oxana^{1*}

`tsidulko.ox@gmail.com`

¹Novosibirsk, Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences

²Novosibirsk, Novosibirsk State University

Three important modifications of the well known traveling salesman problem have been considered: m peripatetic salesman problem with different weight functions for each salesman, m cycles cover, and m chains cover problems. In the first problem, given a complete directed or undirected n -vertex graph $G = (V, E)$ with weight functions $w_i : E \rightarrow \mathbf{R}_+$, $i = 1, \dots, m$, and it is necessary to find m edge-disjoint Hamiltonian circuits $H_1, \dots, H_m \subset E$ so as to minimize their total weight:

$$\sum_{i=1}^m w_i(H_i) = \sum_{i=1}^m \sum_{e \in H_i} w_i(e).$$

The second and the third problems consist in finding, given a complete weighted graph G , m cycles (chains) of minimal total weight, such that each vertex in G belongs to exactly one cycle (chain)

All three problems are strong NP-hard, since the traveling salesman problem reduces to each of them. For these problems, approximation polynomial algorithms have been proposed based on the greedy idea. The behavior of these algorithms has been studied under assumption that the weights of the edges of G are independent and identically distributed real numbers [1].

The conditions for the algorithms to be asymptotically optimal have been found in the cases when the weights of the edges of G have uniform distribution in $[a_n, b_n]$, $0 < a_n < b_n$, and exponential distribution with parameter β_n in $[a_n, \infty)$, $a_n > 0$.

This research is funded by the Russian Foundation for Basic Research, grant 15-01-00976 and 13-07-00070.

- [1] Gimadi, E., Istomin A., Rykov I., and Tsidulko O. 2014. Probabilistic analysis for an approximation algorithm solving m peripatetic salesman problem on random inputs unbounded from above. *Tr. IMM UrB RAS* 20-2:88–98. (In Russian.) <http://mi.mathnet.ru/timm1061>.

Приближенные методы минимизации булевых функций в специальных классах

Гранин Сергей Сергеевич^{1*}

ssgranin@gmail.com

Максимов Юрий Владимирович^{2,3,4}

yurymaximov@iitp.ru

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Институт проблем передачи информации РАН

³Москва, Московский физико-технический институт (ГУ),
Лаборатория структурных методов анализа данных
в предсказательном моделировании (ПреМоЛаб)

⁴Москва, Национальный исследовательский университет «Высшая
школа экономики»

Рассматриваемая задача посвящена приближенным методам построения простых (кратчайших и минимальных) дизъюнктивных нормальных форм булевых функций, принимающих нулевое значение в относительно небольшом числе точек. Наиболее интересным представляется случай, в котором число указанных точек ограничено полиномом от размерности пространства переменных.

Данные функции играют достаточно важную роль в решении задач распознавания логическими методами, при синтезе функциональных схем, в задачах тестирования дискретных устройств.

Ключевым вкладом являются новые границы на сложность (длину, ранг) булевых функции с ограниченным числом нулей, равно как и эффективные методы их построения. Представлены также некоторые алгоритмы для решения задачи минимизации частично заданных функций. Близкие результаты представлены в работе [1].

Работа поддержана грантами РФФИ 14-07-31277 мол_а и 15-07-09121.

- [1] *Гранин С. С., Максимов Ю. В.* Сложность дизъюнктивных нормальных форм и полу-эффект Шеннона в некоторых подклассах булевых функций // Автоматика и телемеханика, 2015. <http://arxiv.org/pdf/1501.03444v1.pdf>.

Approximate minimization of disjunctive normal forms of Boolean functions from special classes

Granin Sergey^{1*}

ssgranin@gmail.com

Maximov Yury^{2,3,4}

yurymaximov@iitp.ru

¹Moscow, Moscow Institute of Physics and Technology (State University)

²Moscow, Institute for Information Transmission Problems
of the Russian Academy of Sciences

³Moscow, Moscow Institute of Physics and Technology (State University),
Laboratory of Structural Methods in Predictive Modeling (PreMoLab)

⁴Moscow, National Research University “Higher School of Economics”

The considered problem is devoted to accurate approximation of simplest (shortest and minimal) disjunctive normal forms of Boolean functions that attain False value in a relatively small number of points. The class of functions with number of False points polynomially bounded with respect to the dimension of input space is of primal interest.

Such kind of functions plays an important role for the purposes of logical analysis of data, algebraic methods of machine learning, testing, and other areas.

The key contribution is a new state-of-the-art methods that give upper bounds for approximation of shortest and minimal disjunctive normal forms of Boolean functions with relatively small number of zeros. Some results for disjunctive normal form minimization of partially defined Boolean are presented as well.

The main tool of the proposed approach is Fourier analysis over Boolean hypercube. Tightly connected results are suggested earlier in [1].

The work is supported by the Russian Foundation for Basic Research, grants 14-07-31277 mol_a and 15-07-09121.

- [1] Granin, S., and Maximov Yu. 2015. Average case complexity of DNFs and Shannon semi-effect for narrow subclasses of boolean functions. arXiv:1501.03444.

Построение хабов для рынка электроэнергии

*Еремеев Антон Валентинович** eremeev@fim.oscsbras.ru
Омск, Омский филиал Института математики им. С. Л. Соболева
СО РАН

Рассматривается задача построения торговых хабов для рынка электроэнергии с маргинальным узловым ценообразованием. Задано число искомых хабов и известны значения цен электроэнергии для каждого узла сети и каждого субъекта рынка за некоторый достаточно продолжительный период. Требуется выбрать кластеры узлов сети (хабы) и прикрепить к ним субъекты рынка с целью минимизации суммарного взвешенного среднеквадратического отклонения узловых цен субъектов от индексов хабов. В качестве ограничений задано условие достаточно низкой концентрации рынка в каждом из хабов, что измеряется индексом Херфиндала–Хиршмана, а также нижняя граница на число узлов хаба.

Установлено, что рассматриваемая задача является NP трудной в сильном смысле и для нее не существует вполне полиномиальной аппроксимационной схемы (FPTAS), если $P \neq NP$. Ввиду вычислительной сложности задачи, для ее решения предложены эволюционные алгоритмы: генетический алгоритм и гибридная эвристика локального поиска. Проведено тестирование и сравнение алгоритмов на реальных данных. Исследованы свойства «ландшафта» локальных оптимумов с использованием многократного локального спуска, что позволило объяснить особенности работы эволюционных алгоритмов. С практической точки зрения разработанные алгоритмы могут служить полезным средством поддержки принятия решений при построении торговых хабов для оптовых рынков электроэнергии [1].

Работа поддержана грантом РФФИ № 15-11-10009.

- [1] *Borisovsky P., Eremeev A., Grinkevich E., Klokov S., Kosarev N.* Trading hubs construction in electricity markets using evolutionary algorithms // *Pattern Recognition Image Anal.*, 2014. Vol. 24. Iss. 2. P. 270–282. http://dx.doi.org/10.1007/978-3-540-88965-6_3.

Trading Hubs Construction for Electricity Markets

*Eremeev Anton**

eremeev@ofim.oscsbras.ru

Omsk, Omsk Branch of Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences

The trading hubs construction problem for electricity markets under locational marginal prices is considered. Given historical prices for all nodes of the electricity grid and for all market participants over a sufficiently long period of time, the problem is to choose a required number of node clusters (hubs) and to assign market participants to hubs so as to minimize the deviation of hub prices from the prices of participants. The problem is formulated under constraints of sufficiently low market concentration, measured in terms of Herfindahl–Hirschmann index, and lower bounds on hub sizes.

The problem is shown to be strongly NP-hard and it does not admit a fully polynomial approximation scheme (FPTAS) if $P \neq NP$. In view of problem complexity, two evolutionary algorithms are proposed: a genetic algorithm and a hybrid local search heuristic. The algorithms are tested and compared on the real-life data. The properties of the fitness landscapes are analyzed using multiple restarts of the local search and the behavior of the evolutionary algorithms is explained on the basis of this analysis. From the practical point of view, the developed algorithms proved to be a useful decision-support tool in trading hubs construction for the electricity wholesale markets [1].

This research is funded by the Russian Science Foundation, grant 15-11-10009.

- [1] Borisovsky, P., Eremeev A., Grinkevich E., Klokov S., and Kosarev N. 2014. Trading hubs construction in electricity markets using evolutionary algorithms. *Pattern Recognition Image Anal.* 24(2):270–282. http://dx.doi.org/10.1007/978-3-540-88965-6_3.

Отбор объектов и признаков для монотонных классификаторов

*Зухба Анастасия Викторовна**

a__l@mail.ru

Москва, Московский физико-технический институт (ГУ)

Некоторые прикладные задачи классификации предполагают наличие монотонных зависимостей между признаками и вероятностью одного из классов. В таких задачах часто используются линейные классификаторы. Расширение модели классификации до семейства монотонных функций повышает точность аппроксимации обучающей выборки, но также повышает риск переобучения, поэтому для монотонных классификаторов важна оптимизация сложности модели путем отбора признаков. Кроме того, построение монотонного классификатора предполагает монотонизацию выборки путем отбора подмножества опорных объектов, удовлетворяющих условию монотонности. Таким образом, при построении монотонных классификаторов возникает задача одновременного отбора объектов и признаков. Для ее решения, как правило, используют жадные эвристические алгоритмы, однако вопросы вычислительной сложности задачи построения монотонных классификаторов относительно мало исследованы.

В данной работе рассматриваются возможные постановки задачи монотонизации выборки как задачи дискретной оптимизации. Предлагается систематизация задач одновременного отбора объектов и признаков для монотонизации выборки. Для каждой постановки задачи монотонизации устанавливается статус ее вычислительной сложности. Доказывается полиномиальная сводимость к задачам монотонизации таких известных NP-полных задач, как задача о рюкзаке, задача о биклике, задача о минимальном покрытии множества подмножествами. Для постановок, разрешимых за полиномиальное время, предлагается точный алгоритм решения [1].

Работа поддержана РФФИ, грант № 14-07-31240.

- [1] *Зухба А. В.* Вычислительная сложность отбора объектов и признаков для задач с ограничениями монотонности // Математическая биология и биоинформатика, 2015 (в печати).

Feature and prototype selection for isotonic classifier

Zukhba Anastasia*

a_l@mail.ru

Moscow, Moscow Institute of Physics and Technology (State University)

In many applications of two-class classification, it is known beforehand that the class label should be increasing (or decreasing) in the features. In such situations, linear classifiers are usually used. Expanding classification model to the family of isotonic functions reduces the approximation error on the training sample but also increases the risk of overfitting. Therefore, the optimization of the model complexity via feature selection is an important issue for isotonic classifiers. In addition, learning isotonic classifier assumes a monotonicity of the training set by sampling a subset of prototype objects satisfying monotonicity constraints. Thus, learning isotonic classifier requires simultaneous selection of objects and features. Greedy heuristic algorithms are typically used to solve this problem. Nevertheless, the computational complexity of the corresponding discrete optimization problems still remains little studied.

In this work, a systematization of monotonicity problem statements is proposed as discrete optimization problems. For each problem, its computational complexity is estimated. It is shown that well-known NP-complete problems, such as the knapsack problem, and the biclique problem, the minimal covering set problem are polynomially reducible to different variants of the monotonicity problem. An effective algorithm for polynomially solvable variants of the monotonicity problem is presented [1].

This research is funded by the Russian Foundation for Basic Research, grant 14-07-31240.

- [1] Zukhba, A. 2015 (in press). Computational complexity of prototype and feature selection for isotonic classification problems *Math. Biol. Bioinf.*

Квадратичные евклидовы задачи 2-кластеризации: сложность и эффективные алгоритмы с оценками точности для их решения

Кельманов Александр Васильевич^{1,2*} kelm@math.nsc.ru

¹Новосибирск, Институт математики им. С. Л. Соболева СО РАН

²Новосибирск, Новосибирский государственный университет

Рассмотрены несколько квадратичных евклидовых задач 2-кластеризации, которые имеют приложения в анализе данных, распознавании образов, статистике, комбинаторной геометрии, теории приближения. Задачи актуальны для многих естественно-научных и технических приложений. Цель статьи — обзор новых (неизученных), известных (но слабоизученных) задач, а также результатов о вычислительной сложности этих задач и эффективных алгоритмах с оценками точности для их решения.

Ниже приведен список рассматриваемых задач.

Задача 1. *Minimum sum-of-squares 2-clustering with given center of one cluster.*

Задача 2. *Minimum sum-of-squares 2-clustering problem on sequence with given center of one cluster.*

Задача 3. *Quadratic Euclidean Max-Cut.*

Задача 4. *Quadratic Euclidean Min-Sum all-pairs 2-clustering.*

Задача 5. *Euclidean balanced variance-based 2-clustering.*

Задача 6. *Euclidean balanced variance-based 2-clustering with given center of one cluster.*

Основное внимание уделено задачам разбиения конечных множеств и последовательностей точек евклидова пространства на два кластера при заданном (желаемом) центре одного из кластеров [1]. Желаемый центр задается на входе в некоторой точке, без ограничения общности — в начале координат, а второй центр неизвестен. Анализируются два варианта задач, в которых мощности кластеров либо являются частью входа, либо неизвестны.

Работа поддержана грантом РФФИ № 15-01-00462.

- [1] Кельманов А. В. Квадратичные евклидовы задачи би-разбиения множества и последовательности: сложность и алгоритмы с оценками качества решения // Мат-лы VI Междунар. конф. «Проблемы оптимизации и экономические приложения». — Омск, 2015. С. 36–41.

Quadratic Euclidian 2-clustering problems: Complexity and efficient algorithms with performance guarantees for their solutions

Kel'manov Alexander^{1,2*}

kelm@math.nsc.ru

¹Novosibirsk, Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences

²Novosibirsk, Novosibirsk State University

Some quadratic Euclidean 2-clustering problems induced by actual issues in data analysis, pattern recognition, statistics, computational geometry, and approximation theory have been considered. These problems are also important in a wide range of natural-science and engineering applications. The purpose of the paper is to overview new (previously unstudied), known (poorly studied), and recent results on computational complexity of these problems and on efficient algorithms with performance guarantees for their solutions.

Below is the list of the problems considered.

Problem 1. *Minimum sum-of-squares 2-clustering with given center of one cluster.*

Problem 2. *Minimum sum-of-squares 2-clustering problem on sequence with given center of one cluster.*

Problem 3. *Quadratic Euclidean Max-Cut.*

Problem 4. *Quadratic Euclidean Min-Sum all-pairs 2-clustering.*

Problem 5. *Euclidean balanced variance-based 2-clustering.*

Problem 6. *Euclidean balanced variance-based 2-clustering with given center of one cluster.*

The focus is on the 2-clustering problems (for the finite set and finite sequence of points in the Euclidean space) with the given center of one cluster [1]. So, the desired center of one of the clusters is given (without loss of generality at 0) as an input, while the center of the second cluster is unknown (a variable for optimizing). Two variants of the problems are analyzed, where the cardinalities of the clusters are either the parts of input or unknown.

This research is funded by the Russian Foundation for Basic Research, grant 15-01-00462.

- [1] *Kel'manov, A.* 2015. Quadratic Euclidian bi-clustering problems: Complexity and efficient algorithms with performance guarantees. *6th Conference (International) "Optimization Problems and their Economical Applications" Proceedings*. Omsk. 36–41.

Точный псевдополиномиальный алгоритм для одной квадратичной евклидовой задачи сбалансированной 2-кластеризации

Кельманов Александр Васильевич^{1,2} kelm@math.nsc.ru

Моткова Анна Владимировна^{2*} anitam@mail.ru

¹Новосибирск, Институт математики им. С. Л. Соболева СО РАН

²Новосибирск, Новосибирский государственный университет

Рассматривается следующая NP-трудная в сильном смысле

Задача (Euclidean balanced variance-based 2-clustering with given center of one cluster and cluster cardinalities). *Дано:* множество $\mathcal{Y} = \{y_1, \dots, y_N\}$ точек из \mathbb{R}^q и натуральное число M . Найти разбиение множества \mathcal{Y} на кластеры \mathcal{C} и $\mathcal{Y} \setminus \mathcal{C}$ такие, что

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

где $\bar{y}(\mathcal{C}) = (1/|\mathcal{C}|) \sum_{y \in \mathcal{C}} y$ — центр масс \mathcal{C} , при ограничении $|\mathcal{C}| = M$.

Задача актуальна, в частности, при решении проблем анализа и интерпретации данных.

В настоящей работе построен точный псевдополиномиальный алгоритм, который находит оптимальное решение задачи для случая целочисленных координат точек входного множества и фиксированной размерности пространства [1]. Время работы алгоритма есть величина $\mathcal{O}(N(MD)^q)$, где D — максимальное абсолютное значение координат точек входного множества.

Работа поддержана грантами РФФИ № 13-07-00070 и № 15-01-00462.

[1] *Кельманов А.В., Моткова А.В.* Точный псевдополиномиальный алгоритм для одной квадратичной евклидовой задачи сбалансированной 2-кластеризации // *Дискретный анализ и исследование операций*, 2015 (в печати).

An exact pseudopolynomial algorithm for an Euclidean balanced variance-based 2-clustering problem

Kel'manov Alexander^{1,2}

kelm@math.nsc.ru

*Motkova Anna*²★

anitamo@mail.ru

¹Novosibirsk, Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences

²Novosibirsk, Novosibirsk State University

The authors consider the following strongly NP-hard

Problem (Euclidean balanced variance-based 2-clustering with given center of one cluster and cluster cardinalities). *Given:* a set $\mathcal{Y} = \{y_1, \dots, y_N\}$ of points from \mathbb{R}^q and a positive integer number M . *Find:* a partition of \mathcal{Y} into clusters \mathcal{C} and $\mathcal{Y} \setminus \mathcal{C}$ such that

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

where $\bar{y}(\mathcal{C}) = (1/|\mathcal{C}|) \sum_{y \in \mathcal{C}} y$ is the centroid of \mathcal{C} , under constrain $|\mathcal{C}| = M$.

This problem is actual, in particular, in data analysis.

A pseudopolynomial algorithm which finds an optimal solution for the case of integer components of the points in the input set and fixed space dimension is presented [1]. The running time of the algorithm is $\mathcal{O}(N(MD)^q)$ where D is the maximum absolute coordinate value of the points in the input set.

This research is funded by the Russian Foundation for Basic Research, grants 13-07-00070 and 15-01-00462.

- [1] Kel'manov, A. V., and Motkova A. V. 2015 (in press). Tochnyy psevdopolynomial'nyy algoritm dlya odnoy kvadrachnoy evklidovoy zadachi sbalansirovannoy 2-klasterizatsii An exact pseudopolynomial algorithm for a quadratic Euclidean balanced 2-clustering problem // Diskretnyy analiz i issledovanie operatsiy [Discrete Analysis and Operations Research].

О сложности одной квадратичной евклидовой задачи сбалансированной 2-кластеризации

Кельманов Александр Васильевич^{1,2} kelm@math.nsc.ru
Пяткин Артем Валерьевич^{2*} artem@math.nsc.ru

¹Новосибирск, Институт математики им. С. Л. Соболева СО РАН

²Новосибирск, Новосибирский государственный университет

Рассматривается задача разбиения конечного множества точек евклидова пространства на два кластера по критерию минимума суммы весов кластеров, умноженных на их мощности. Под весом кластера понимается сумма квадратов расстояний от элементов кластера до его центра. Центр одного кластера задан, а центр другого — неизвестен, является геометрическим центром (центроидом) и определяется как среднее значение элементов этого кластера. Задача формулируется следующим образом.

Задача Euclidean Balanced Variance-based 2-Clustering with Given Center of one cluster. *Дано:* множество $\mathcal{Y} = \{y_1, \dots, y_N\}$ точек из \mathbb{R}^q . *Найти:* разбиение множества \mathcal{Y} на два непустых подмножества \mathcal{X} и \mathcal{Z} таких, что

$$|\mathcal{X}| \sum_{x \in \mathcal{X}} \|x - \bar{x}(\mathcal{X})\|^2 + |\mathcal{Z}| \sum_{z \in \mathcal{Z}} \|z\|^2 \longrightarrow \min,$$

где $\bar{x}(\mathcal{X}) = (1/|\mathcal{X}|) \sum_{x \in \mathcal{X}} x$ — центроид подмножества \mathcal{X} .

Эта задача индуцируется, в частности, проблемами анализа данных и распознавания образов комбинаторной геометрии, статистики. Вопрос о статусе сложности задачи был открытым.

В настоящей работе доказано [1], что (1) задача NP-трудна в сильном смысле и (2) для нее не существует полностью полиномиальной приближенной схемы (FPTAS), если $P \neq NP$. Для доказательства факта труднорешаемости задачи построено полиномиальное сведение к ней NP-трудной в сильном смысле задачи *Minimum Bisection* для кубических графов.

Работа поддержана грантами РФФИ №№ 15-01-00462, 15-01-00976 и 13-07-00070.

[1] *Кельманов А. В., Пяткин А. В.* NP-трудность некоторых квадратичных евклидовых задач 2-кластеризации // Докл. РАН, 2015 (в печати).

On the complexity of Quadratic Euclidean Balanced 2-Clustering problem

Kel'manov Alexander^{1,2}

kelm@math.nsc.ru

Pyatkin Artem^{2*}

artem@math.nsc.ru

¹Novosibirsk, Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences

²Novosibirsk, Novosibirsk State University

The following problem has been considered. Given a finite set of points from Euclidean space, find a partition of this set into two clusters minimizing the sum of weights of the clusters multiplied by their cardinalities. By the weight of the cluster, the sum of squared distances from the elements of the cluster to its center has been meant. The center of one cluster is given while for the second cluster, it is unknown (so, it is estimated by the centroid that is the mean value of the elements of the cluster).

Problem (Euclidean Balanced Variance-based 2-Clustering with Given Center of one cluster). *Given:* a set $\mathcal{Y} = \{y_1, \dots, y_N\}$ of points from \mathbb{R}^q . *Find:* a partition of set \mathcal{Y} into two subsets \mathcal{X} and \mathcal{Z} such that

$$|\mathcal{X}| \sum_{x \in \mathcal{X}} \|x - \bar{x}(\mathcal{X})\|^2 + |\mathcal{Z}| \sum_{z \in \mathcal{Z}} \|z\|^2 \rightarrow \min,$$

where $\bar{x}(\mathcal{X}) = (1/|\mathcal{X}|) \sum_{x \in \mathcal{X}} x$ is the centroid of subset \mathcal{X} .

This problem is actual, in particular, for solving the problems in data analysis. The complexity of this problem was unexplored.

It has been proven [1] that this problem (i) is strongly NP-hard and (ii) do not admit FPTAS (fully polynomial-time approximation scheme), unless $P = NP$. The proof is based on polynomial-time reduction from the strongly NP-hard *Minimum Bisection* problem on cubic graphs.

This research is funded by the Russian Foundation for Basic Research, grants 15-01-00462, 15-01-00976, and 13-07-00070.

[1] Kel'manov, A., and Pyatkin A. 2015 (in press). NP-hardness of some Quadratic Euclidean 2-clustering problems. *Dokl. Math.*

Полностью полиномиальная приближенная схема для одной задачи 2-кластеризации последовательности

Кельманов Александр Васильевич^{1,2} kelm@math.nsc.ru

Хамидуллин Сергей Асгадулович^{2*} kham@math.nsc.ru

*Хандеев Владимир Ильич*² khandeev@math.nsc.ru

¹Новосибирск, Новосибирский государственный университет

²Новосибирск, Институт математики им. С. Л. Соболева СО РАН

Рассматривается следующая NP-трудная в сильном смысле

Задача (Minimum sum-of-squares 2-clustering problem on sequence with given center of one cluster and cluster cardinalities).

Дано: последовательность $\mathcal{Y} = (y_1, \dots, y_N)$ точек из \mathbb{R}^q , натуральные числа T_{\min} , T_{\max} и $M > 1$. *Найти:* подмножество $\mathcal{M} = \{n_1, \dots, n_M\} \subseteq \mathcal{N}$ номеров элементов последовательности \mathcal{Y} такое, что

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

где $\bar{y}(\mathcal{M}) = (1/|\mathcal{M}|) \sum_{i \in \mathcal{M}} y_i$ при ограничениях

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

на элементы набора (n_1, \dots, n_M) .

Рассматриваемая задача актуальна, в частности, при решении проблем помехоустойчивой обработки временных рядов.

В настоящей работе установлено [1], что если $P \neq NP$, то для задачи не существует полностью полиномиальной приближенной схемы (FPTAS). Такая схема построена для случая фиксированной размерности пространства. Трудоемкость алгоритма, реализующего схему FPTAS, равна $\mathcal{O}(N^4(1/\varepsilon)^{q/2})$, где ε — гарантированная относительная погрешность.

Работа поддержана грантами РФФИ № 13-07-00070 и № 15-01-00462.

[1] *Кельманов А. В., Хамидуллин С. А., Хандеев В. И.* FPTAS для одной задачи 2-кластеризации последовательности // Дискретный анализ и исследование операций, 2015 (в печати).

Fully polynomial-time approximation scheme for a sequence 2-clustering problem

Kel'manov Alexander^{1,2}

kelm@math.nsc.ru

Khamidullin Sergey^{2*}

kham@math.nsc.ru

*Khandeev Vladimir*²

khandeev@math.nsc.ru

¹Novosibirsk, Novosibirsk State University

²Novosibirsk, Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences

The authors consider the following strongly NP-hard

Problem (Minimum sum-of-squares 2-clustering problem on sequence with given center of one cluster and cluster cardinalities). *Given:* a sequence $\mathcal{Y} = (y_1, \dots, y_N)$ of points from \mathbb{R}^q and some positive integer numbers T_{\min} , T_{\max} , and M . *Find:* a subset $\mathcal{M} = \{n_1, \dots, n_M\} \subseteq \mathcal{N} = \{1, \dots, N\}$ such that

$$\sum_{j \in \mathcal{M}} \|y_j - \bar{y}(\mathcal{M})\|^2 + \sum_{i \in \mathcal{N} \setminus \mathcal{M}} \|y_i\|^2 \rightarrow \min$$

where $\bar{y}(\mathcal{M}) = (1/|\mathcal{M}|) \sum_{i \in \mathcal{M}} y_i$, under constraints

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

on the elements of set (n_1, \dots, n_M) .

The considered problem is actual, in particular, in noise-proof analysis of time series.

It has been proven [1] that unless $P = NP$, in the general case of the problem, there is no fully polynomial-time approximation scheme (FPTAS). In addition, such a scheme is presented for the case of fixed space dimension. The running time of the algorithm is $\mathcal{O}(N^4(1/\varepsilon)^{q/2})$ where ε is the arbitrary relative error.

This research is funded by the Russian Foundation for Basic Research, grants 13-07-00070 and 15-01-00462.

[1] Kel'manov, A., Khandeev V., and Khamidullin S. 2015 (in press). FPTAS for a sequence 2-clustering problem. *J. Appl. Indust. Math.*

Полностью полиномиальная приближенная схема для одной квадратичной евклидовой задачи 2-кластеризации

Кельманов Александр Васильевич^{1,2} kelm@math.nsc.ru

*Хандеев Владимир Ильич*²* khandeev@math.nsc.ru

¹Новосибирск, Новосибирский государственный университет

²Новосибирск, Институт математики им. С. Л. Соболева СО РАН

Рассматривается следующая NP-трудная в сильном смысле

Задача (Minimum sum-of-squares 2-clustering problem with given center of one cluster and cluster cardinalities). *Дано:* множество $\mathcal{Y} = \{y_1, \dots, y_N\}$ точек из \mathbb{R}^q и натуральное число M . *Найти:* разбиение множества \mathcal{Y} на два кластера \mathcal{C} и $\mathcal{Y} \setminus \mathcal{C}$ такое, что

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

где $\bar{y}(\mathcal{C}) = (1/|\mathcal{C}|) \sum_{y \in \mathcal{C}} y$ — центроид кластера \mathcal{C} , при ограничении $|\mathcal{C}| = M$.

Рассматриваемая задача актуальна, в частности, в анализе данных и распознавании образов, математической статистике, компьютерной геометрии.

В настоящей работе установлено [1], что если $P \neq NP$, то для задачи не существует полностью полиномиальной приближенной схемы (FPTAS). Такая схема построена для случая фиксированной размерности пространства. Трудоемкость алгоритма, реализующего схему FPTAS, равна $\mathcal{O}(N^3(1/\varepsilon)^{q/2})$, где ε — гарантированная относительная погрешность.

Работа поддержана грантами РФФИ № 13-07-00070 и № 15-01-00462.

- [1] *Kel'manov A. V., Khandeev V. I.* FPTAS for a special case of a quadratic Euclidean 2-clustering problem // *Comput. Math. Math. Phys.*, 2015 (in press).

Fully polynomial-time approximation scheme for a quadratic Euclidean 2-clustering problem

Kel'manov Alexander^{1,2}

kelm@math.nsc.ru

Khandeev Vladimir^{2*}

khandeev@math.nsc.ru

¹Novosibirsk, Novosibirsk State University

²Novosibirsk, Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences

The authors consider the following strongly NP-hard

Problem (Minimum sum-of-squares 2-clustering problem with given center of one cluster and cluster cardinalities). *Given:* a set $\mathcal{Y} = \{y_1, \dots, y_N\}$ of points from \mathbb{R}^q and a positive integer number M . *Find:* a partition of \mathcal{Y} into clusters \mathcal{C} and $\mathcal{Y} \setminus \mathcal{C}$ such that

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

where $\bar{y}(\mathcal{C}) = (1/|\mathcal{C}|) \sum_{y \in \mathcal{C}} y$ is the centroid of \mathcal{C} , under constraint $|\mathcal{C}| = M$.

The considered problem is actual, in particular, in data analysis and pattern recognition, computer geometry, and mathematical statistics.

It has been proven that unless $P = NP$, in the general case of the problem, there is no fully polynomial-time approximation scheme (FPTAS) [1]. In addition, such a scheme is presented for the case of fixed space dimension. The running time of the algorithm is $\mathcal{O}(N^3(1/\varepsilon)^{q/2})$ where ε is the arbitrary relative error.

This research is funded by the Russian Foundation for Basic Research, grants 13-07-00070 and 15-01-00462.

[1] Kel'manov, A., and Khandeev V. 2015 (in press). FPTAS for a special case of a quadratic Euclidean 2-clustering problem. *Comput. Math. Math. Phys.*

Модификации локального элиминационного алгоритма для эффективного решения больших разреженных задач дискретной оптимизации

Лемтюжникова Дарья Владимировна darabbt@gmail.com

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

Один из подходов к решению разреженных задач целочисленного линейного программирования основан на выделении блочно-древовидной (БД) структуры в матрице ограничений. В частности, модифицированный метод декомпозиции Финкельштейна позволяет выделять сбалансированную структуру. Затем производится декомпозиция задачи, например с помощью локального элиминационного алгоритма (ЛЭА), предложенного О. А. Щербиной. Локальный элиминационный алгоритм сводит исходную задачу большой размерности к последовательности задач меньших размерностей, связанных с отдельными блоками БД-структуры. Проблема данного подхода заключается в том, что для многих задач большой размерности поиск точного решения таким способом остается чрезвычайно трудоемким. Целью данной работы является расширение класса задач, решаемых с помощью ЛЭА.

Были проведены исследования различных методов сокращения перебора в ЛЭА как точных, так и приближенных. Лексикографический поиск оказался наиболее эффективной эвристикой для решения NP-трудной задачи определения оптимального порядка элиминации. Препроцессинг показал свою эффективность для предварительной обработки задач большой размерности. Использование постоптимального анализа сокращает время решения задач для больших пакетов типовых задач. Релаксации позволяют получать решения для многих задач большой размерности значительно быстрее. Распараллеливание существенно уменьшает время нахождения решений для задач с большими сепараторами. Локальный элиминационный алгоритм с оракулом оказался эффективным для быстрого получения приближенного решения [1].

- [1] *Лемтюжникова Д. В.* Параллельное представление локального элиминационного алгоритма для ускорения решения разреженных задач дискретной оптимизации // Компьютерные исследования и моделирование, 2015. Т. 7. № 3. С. 699–705.

Modification of local elimination algorithm for sparse large-scale problems of discrete optimization

*Lemtyuzhnikova Darya**

darabbt@gmail.com

Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

An approach to solving sparse integer linear programming problems is considered. It is based on the allocation of block-tree structure in the matrix of restrictions. The modified Finkelstein decomposition is used to select a well balanced block-tree structure. Thereafter, the local elimination algorithm proposed by O. A. Shcherbina is used for the decomposition of the problem. The local elimination algorithm reduces the initial large-scale problem to a sequence of low-dimensional problems associated with individual blocks of the block-tree structure. Reducing the time complexity of the exact solution still remains a challenging problem for many large-scale applications. The goal of this work is to extend the class of problems effectively solved with the local elimination algorithm.

Various methods of reducing the running time of the local elimination algorithm, both exact and approximate, have been investigated. It is shown that lexicographic search is the most efficient heuristic for the NP-hard task of determining the optimal elimination order. The preprocessing is effective for large-scale problems. The postoptimal analysis reduces the total running time for large packages of typical tasks. Relaxation reaches a solution much faster for many large-scale problems. Parallelization reduces the running time for problems with large separators. The local elimination algorithm with oracle is effective for finding approximate solution quickly [1].

- [1] Lemtyuzhnikova, D. 2015. Parallel representation of local elimination algorithm for accelerating the solving sparse discrete optimization problems. *Computer Research Modeling* 7(3):699–705.

Оптимистические оценки риска в задачах многоклассовой классификации

Максимов Юрий Владимирович^{1,2,3} yurymaximov@iitp.ru
Решетова Дарья Германовна^{2*} reshetova@phystech.edu

¹Москва, Институт проблем передачи информации РАН

²Москва, Московский физико-технический институт (ГУ),
Лаборатория структурных методов анализа данных
в предсказательном моделировании

³Москва, Национальный исследовательский университет «Высшая школа экономики»

Рассматривается задача многоклассовой классификации с дополнительными предположениями о разделимости разных данных в однородные по классам регионы. Основная идея предлагаемого алгоритма суть локализация сложности задачи с большим числом классов на основе анализа схожести признаков описаний данных.

Представлены новые верхние и нижние границы на обобщающую способность многоклассовых классификаторов. Также обсуждаются вопросы уточнения полученных границ для случая обучения функциональных классов малой сложности.

На основе предлагаемых оценок предложены новые многошаговые алгоритмы многоклассовой классификации и продемонстрирована их эффективность на различных данных из репозитория ImageNet и UCI [1].

Работа поддержана РФФИ, грант 15-07-09121_а.

[1] *Maximov Yu., Reshetova D.* Tight risk bounds for multi-class margin classifiers. <http://arxiv.org/abs/1507.03040>.

Optimistic risk bounds in multiclass classification problems

Maximov Yury^{1,2,3}

yurymaximov@iitp.ru

Reshetova Daria^{2*}

reshetova@phystech.edu

¹Moscow, Institute for Information Transmission Problems
of the Russian Academy of Sciences

²Moscow, Institute of Physics and Technology (State University),
Laboratory of Structural Methods in Predictive Modeling (PreMoLab)

³Moscow, National Research University “Higher School of Economics”

A problem of multiclass classification under plausible assumptions on data separability into label homogeneous regions is considered. The main idea of the algorithm is to localize a multiclass complexities based on the analysis of data similarity.

Upper and lower Rademacher complexity based bounds on the multiclass risk bounds are provided. The properties of the class functions to be learned are discussed which allow to reduce the risk bounds dependence in the number of classes as well.

It seems that the bounds matches with state-of-the-art results and can be generalized on transductive and semisupervised setting.

The efficiency of the method is illustrated on various data from ImageNet and UCI datasets [1].

The work is supported by the Russian Foundation for Basic Research, grant 15-07-09121_a.

[1] Maximov, Yu., and Reshetova D. Tight risk bounds for multi-class margin classifiers. <http://arxiv.org/abs/1507.03040>

Существование равновесий в двухуровневых задачах ценообразования и размещения хабов

Плясунов Александр Владимирович^{1*} apljas@math.nsc.ru

*Кочетов Юрий Андреевич*¹ jkochet@math.nsc.ru

*Чвокич Дмитрий*² dimitriye.chwokitch@yahoo.com

¹Новосибирск, Институт математики им. С.Л. Соболева СО РАН

²Баня-Лука, Университет Баня-Лука

В задачах размещения хабы служат в качестве пунктов, в которых накапливаются и перераспределяются потоки, например, пассажиров, грузов, почты, данных. Эффект от введения хабов обусловлен с одной стороны удалением части маршрутов, по которым происходит передача потока, а с другой стороны экономией за счет роста объемов перевозок между хабами.

В работе предлагается новая двухуровневая нелинейная частично-целочисленная модель размещения хабов, связанных с ними радиальных сетей, и выбора цен за обслуживание клиентов. Лидер первым выбирает свои хабы, конструирует топологию своих радиальных сетей и устанавливает цены. Затем свой выбор делает последователь. Для каждой пары (пункт_генерации, пункт_назначения) и в сети лидера, и в сети последователя найдется свой маршрут передачи потока. Распределение потоков между маршрутами описывается специальной моделью (logit model).

Показано [1], что в предлагаемой модели существует равновесие Штакельберга для любого типа взаимодействия игроков: как кооперативного, так и некооперативного. Установлено, что если хабы и сети игроками уже выбраны и игроки участвуют в ценовой войне по типу игры Бертрана, то существует единственное и конечное равновесие Нэша для любого типа взаимодействия игроков. Из этого результата следует, что если рассматривать конкуренцию между игроками при выборе хабов с последующей ценовой войной, то для лидера существует решение, приносящее ему максимальный доход для любого типа взаимодействия игроков.

Работа поддержана грантом РФФИ № 13-07-00016.

- [1] *Čvokić D.D., Kochetov Yu.A., Plyasunov A.V.* Stackelberg equilibria in the leader-follower hub location and pricing problem // 4OR, Springer, 2015 (in press). Vol. 13.

Existence of equilibria in the bilevel hub location and pricing problem

Plyasunov Alexander^{1*}

apljas@math.nsc.ru

*Kochetov Yury*¹

jkochet@math.nsc.ru

*Čvokić Dimitrije*²

dimitriye.chwokitch@yahoo.com

¹Novosibirsk, Sobolev Institute of Mathematics of the Siberian Branch
of the Russian Academy of Sciences

²Banja Luka, University of Banja Luka

Hub location problem is one of the thriving research areas in location theory with many practical applications. Hubs generally serve as consolidation and dissemination facilities through which flows (e. g., passengers, freight, mail, data) are routed between stipulated origins and destinations. Concept itself provides significant reduction in number of links, called spokes, required to ensure that all demand is going to be satisfied. Other benefits are so-called economies of scale. This means that cost per unit of flow is going to be reduced between the hubs. Generally, usage of transportation resources will be much more efficient.

In this paper, a model is proposed where two competitors, called a leader and a follower, are sequentially creating their hub and spoke networks and setting prices. It is called the Leader–Follower Hub Location and Pricing Problem. For this model, it is shown in [1] that there is Stackelberg equilibrium for altruistic and selfish behavior of the follower. Next, the authors address different scenarios in which competitors are playing Bertrand alike pricing game for already given hub and spoke networks. It is shown that there is a finite pricing Nash equilibrium for this case as well.

This research is funded by the Russian Foundation for Basic Research, grant 13-07-00016.

- [1] Čvokić, D.D., Kochetov Yu.A., and Plyasunov A.V. 2015 (in press). Stackelberg equilibria in the leader–follower hub location and pricing problem. *4OR* 13.

Быстрый инкрементальный метод оптимизации больших сумм функций с суперлинейной скоростью сходимости

Родоманов Антон Олегович anton.rodomanov@gmail.com

*Кропотов Дмитрий Александрович**
dmitry.kropotov@gmail.com

Москва, Московский государственный университет
им. М. В. Ломоносова

Рассматривается задача оптимизации следующего вида:

$$\min_{\mathbf{w} \in \mathbb{R}^D} \left[F(\mathbf{w}) := \frac{1}{N} \sum_{i=1}^N f_i(\mathbf{w}) + \frac{\lambda}{2} \|\mathbf{w}\|_2^2 \right],$$

где $f_i : \mathbb{R}^D \rightarrow \mathbb{R}$, $i = 1, \dots, N$, — всюду определенные выпуклые дважды непрерывно дифференцируемые функции; $\mathbf{w} \in \mathbb{R}^D$ — оптимизируемые переменные; $\lambda > 0$ — заданный коэффициент.

Предполагается, что число слагаемых N является очень большим. В такой ситуации для минимизации функции F обычно используют *инкрементальные* методы оптимизации, стоимость итерации которых не зависит от числа функций N .

В последние годы был предложен целый ряд разнообразных инкрементальных методов. К сожалению, несмотря на свою эффективность на практике в некоторых случаях, все эти методы имеют либо сублинейную, либо линейную скорость сходимости.

В этой работе предлагается новый инкрементальный метод оптимизации, обладающий принципиально более быстрой скоростью сходимости — *суперлинейной*. Результаты экспериментального сравнения на реальных данных показывают, что в ряде задач предложенный метод работает существенно быстрее мировых аналогов [1].

Работа поддержана грантом РФФИ №15-31-20596-мол_а_вед, Microsoft Research, research initiative: Computer vision collaborative research in Russia, Skoltech SDP Initiative, applications A1 and A2.

[1] *Родоманов А. О., Кропотов Д. А.* Быстрый инкрементальный метод оптимизации больших сумм функций с суперлинейной скоростью сходимости // Машинное обучение и анализ данных, 2015 (в печати).

Fast incremental optimization of large sums with superlinear convergence rate

Rodomanov Anton

anton.rodomanov@gmail.com

*Kropotov Dmitry**

dmitry.kropotov@gmail.com

Moscow, Lomonosov Moscow State University

The following optimization problem has been considered:

$$\min_{\mathbf{w} \in \mathbb{R}^D} \left[F(\mathbf{w}) := \frac{1}{N} \sum_{i=1}^N f_i(\mathbf{w}) + \frac{\lambda}{2} \|\mathbf{w}\|_2^2 \right]$$

where $f_i : \mathbb{R}^D \rightarrow \mathbb{R}$, $i = 1, \dots, N$ are the convex twice continuously differentiable functions defined everywhere; $\mathbf{w} \in \mathbb{R}^D$ are the optimization variables; and $\lambda > 0$ is the given coefficient.

Assume that the number of variables N is very large. In this case, one usually uses *incremental* optimization methods, whose iteration cost does not depend on N , for optimizing F .

A lot of various incremental methods have been proposed in recent years. Unfortunately, despite being effective in practice in some cases, all these methods still have either a sublinear or linear rate of convergence.

In this work, the authors propose a new incremental optimization method which has a substantially rate of convergence, namely, superlinear. The results of experimental evaluation on a set of real data problems show that the proposed method works much faster than current state-of-the-art methods [1].

This research is funded by the Russian Foundation for Basic Research, grant 15-31-20596-mol_a_ved, Microsoft Research, research initiative: Computer vision collaborative research in Russia, Skoltech SDP Initiative, applications A1 and A2.

- [1] Rodomanov, A., and Kropotov D. 2015 (in press). Fast incremental optimization of large sums with superlinear convergence rate. *J. Mach. Learn. Data Anal.*

Согласование прогнозов в задачах прогнозирования иерархических временных рядов

Стенина Мария Михайловна^{1,2*} mmedvednikova@gmail.com

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Национальный исследовательский университет «Высшая школа экономики»

Рассматривается задача согласования прогнозов семейства временных рядов, объединенных в иерархическую многоуровневую структуру. Вектором независимых прогнозов называется вектор, все элементы которого получены для каждого временного ряда иерархии независимо от других. Вектором согласованных прогнозов называется вектор прогнозов, элементы которого удовлетворяют структуре иерархии. Решением задачи согласования прогнозов является вектор согласованных прогнозов, удовлетворяющих физическим ограничениям и обеспечивающих значение заданной функции потерь не больше, чем при использовании независимых прогнозов.

Предлагается алгоритм теоретико-игрового согласования прогнозов. Задача согласования прогнозов рассматривается как поиск равновесия Нэша в антагонистической игре, где игрок выбирает вектор согласованных прогнозов, удовлетворяющих физическим ограничениям, а природа — вектор действительных значений временных рядов. Доказывается, что выбор в качестве вектора согласованных прогнозов равновесной стратегии игрока гарантирует, что прогнозы будут удовлетворять физическим ограничениям и требованию к качеству. Алгоритм теоретико-игрового согласования находит ближайший вектор к вектору независимых прогнозов, который удовлетворяет условию согласованности и физическим ограничениям. Поскольку точность независимых прогнозов для различных временных рядов может отличаться, то предлагается модифицированный алгоритм теоретико-игрового согласования, который в ходе процедуры согласования в меньшей степени корректирует наиболее точные независимые прогнозы [1].

Работа поддержана грантом РФФИ № 14-07-31046.

- [1] *Стенина М. М., Стрижов В. В.* Согласование прогнозов при решении задач прогнозирования иерархических временных рядов // Информатика и ее применения, 2015. Т. 9. Вып. 2. С. 75–87.

Forecasts reconciliation for hierarchical time series forecasting problem

Stenina Mariya^{1,2*}

mmedvednikova@gmail.com

¹Moscow, Moscow Institute of Physics and Technology (State University)

²Moscow, National Research University “Higher School of Economics”

The forecasts reconciliation problem is investigated. The time series are integrated into the hierarchical multilevel structure. An independent forecasts vector is a vector of forecasts obtained independently for every time series. A reconciled forecasts vector is a vector of forecasts satisfying the hierarchical structure. The solution of forecasts reconciliation problem is a reconciled forecasts vector satisfying the physical constraints and providing value of the loss function no greater than one with independent forecasts.

The game-theoretically optimal reconciliation algorithm is proposed. The forecasts reconciliation problem is seen as Nash equilibrium search in the zero-sum game. The player chooses a reconciled forecasts vector satisfying the physical constraints. And the nature chooses a vector of real values of time series. It is proved that the equilibrium strategy of the player satisfies physical constraints and quality requirements. The game-theoretically optimal reconciliation algorithm finds the reconciled forecasts vector which is the closest one to the independent forecasts vector and satisfies physical constraints. The independent forecasts for different time series have different accuracy. The modified game-theoretically optimal reconciliation algorithm is proposed. It less corrects more accurate independent forecasts [1].

This research is funded by the Russian Foundation for Basic Research, grant 14-07-31046.

- [1] Stenina, M., and Strojov V. 2015. Forecasts reconciliation problem for hierarchical time series forecasting problem. *Informatika i ee Primeneniya* — *Inform. Appl.* 9(2):75–87.

Параметрический синтез робастных систем квазиинвариантного управления методами распознавания образов

*Теклина Лариса Григорьевна**

teklina2010@yandex.ru

Котельников Игорь Вячеславович

neymark@pmk.unn.ru

Нижегородский государственный университет
им. Н. И. Лобачевского

Работа посвящена проблеме синтеза систем управления в условиях действия внешних возмущений. Рассматривается задача параметрического синтеза робастных систем квазиинвариантного управления с регулированием по ошибке управления. Под квазиинвариантностью понимается требуемая малость ошибки управления в установившемся режиме для любого ограниченного по величине внешнего возмущения. Эта задача не является экстремальной и отличается от таковой как по постановке, так и по методам ее решения. Для синтеза таких систем предлагается использовать новый подход, основанный на постановке и решении задачи синтеза в качестве проблемы распознавания образов с активным экспериментом. Методы распознавания работают в пространствах большой размерности и позволяют найти нужную область искомым параметрам с заданной степенью статистической достоверности. С помощью этого подхода можно синтезировать одновременно и робастные, и нехрупкие (или грубые) системы управления. Возможности и особенности применения нового подхода проиллюстрированы на примере двух математических моделей систем управления: управление объектом с эталонной моделью и гашение колебаний высотных сооружений [1].

Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации, проект № 2000.

- [1] *Теклина Л. Г., Котельников И. В.* Анализ и синтез динамических систем методами распознавания образов. — Lap Lambert Academic Publishing, 2015. 136 с.

Parametric synthesis of the robust quasi-invariant control systems using methods of pattern recognition

*Teklina Larisa**

teklina2010@yandex.ru

Kotel'nikov Igor

neymark@pmk.unn.ru

Nizhni Novgorod, Lobachevsky State University of Nizhni Novgorod

The work is devoted to the synthesis of control systems under the action of external disturbances. The problem of parametric synthesis for the robust systems of the quasi-invariant control with regulation in control error is considered. By the quasi-invariance, it is understood the required little control error in the steady state for any bounded external disturbance. This task is not extreme and differs from the same by both the formulation and the methods of its solution. To solve this problem, it is proposed to use a new approach that is based on the formulation and solution of the synthesis problem as a problem of pattern recognition with an active experiment. Methods of pattern recognition work in spaces of large dimension and allow to find the desired region of the unknown parameters with the required degree of statistical reliability. Using this approach, one can synthesize at the same time both the robust and nonfragile (or rough) control systems. The possibilities and peculiarities of the new approach are illustrated by two mathematical models of the control systems: the control object with standard model and vibration damping of high-rise buildings [1].

This work was financially supported by the Ministry of Education and Science of the Russian Federation, project No. 2000.

- [1] Teklina, L., and Kotel'nikov I. 2015. *Analiz i sintez dinamicheskikh sistem metodami raspoznavaniya obrazov* [Analysis and synthesis of dynamical systems using methods of pattern recognition]. Lap Lambert Academic Publishing. 136 p.

Математические морфологии на основе неусложняющих ε -фильтров

*Визильтер Юрий Валентинович**

viz@gosniias.ru

Желтов Сергей Юрьевич

zhl@gosniias.ru

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

Впервые описан принципиально новый подход к алгебраическому описанию теории формы, основанный на понятиях упрощающих модификаторов (УМ-фильтров) и ε -форм. Ключевую роль в предлагаемом морфологическом формализме играют две основные характеристики: *степень модификации* образа фильтром и *сложность* отфильтрованного образа (чем меньше эта сложность, тем сильнее *упрощение* образа фильтром). *Упрощающими модификаторами* предложено называть такие фильтры, которые удовлетворяют следующим двум условиям: неусложнение отфильтрованного образа по отношению к исходному и монотонная зависимость степени модификации от степени упрощения. Как показано в работе, этого достаточно для того, чтобы ряд основных свойств всех известных морфологических фильтров автоматически выполнялся для системы УМ-фильтров. Таким образом, морфология на основе УМ-фильтров является обобщением всех ранее известных схем построения математических морфологий.

Далее, для работы с классами изображений в рамках формализма УМ-фильтров вводятся понятия ε -фильтров и ε -форм. Упрощающий модификатор предложено называть *ε -фильтром*, согласованным с формой заданного эталона a , если степень модификации эталона a данным фильтром не превышает некоторого $\varepsilon \geq 0$. Множество ε -стабильных элементов данного фильтра будем называть *ε -формой* эталона a , если он является ε -фильтром для a . В работе показано, что свойства ε -форм аналогичны свойствам традиционных морфологических классов (0-форм), в частности в задаче сравнения форм по сложности [1].

Работа поддержана грантом РФФИ №14-07-00914-а.

- [1] *Визильтер Ю. В. Желтов С. Ю.* Математические морфологии на основе неусложняющих ε -фильтров // Изв. РАН. ТИСУ, 2015 (в печати).

Mathematical morphologies based on nondecreasing ε -filters

*Vizilter Yuri**

viz@gosniias.ru

Zheltoy Sergey

zhl@gosniias.ru

Moscow, Federal State Unitary Enterprise "State Research Institute of Aviation Systems" (FGUP "GosNIIAS")

The paper describes the new approach to algebraic shape theory description based on the notions of simplifying modifiers (SM-filters) and ε -shapes. Two main characteristics play the key role in the proposed formalism: *modification power* of the filter and *complexity* of the filtered pattern (the less this complexity, the stronger the *simplification power* of the pattern by the filter). It is proposed to call the filter as *simplifying modifier* if it satisfies the following conditions: nondecreasing of filtered pattern relative to the original one and monotonous inverse dependence of modification power with respect to simplification power. It is demonstrated that these properties are enough to reproduce the main properties of all known morphological filters besides the idempotent one. So, the morphology based on SM-filters can be considered as a generalization of all known morphological schemes.

For the operation with image classes, the additional notions of ε -filter and ε -shape are introduced in the framework of SM-filters. Simplifying modifier is called the ε -*filter* for the shape of the given sample image a if the modification power of this filter with regard to a is not greater than $\varepsilon \geq 0$. The set of ε -stable elements of some SM-filter is called an ε -*shape* of a if this filter is an ε -filter of a . It is demonstrated that the properties of ε -shapes are analogous to the properties of traditional morphological classes produced by idempotent filters (0-shapes), in particular, concerning the shape comparison by complexity [1].

This research is funded by the Russian Foundation for Basic Research, grant 14-07-00914-a.

- [1] Vizilter, Yu. V., and Zheltoy S. Yu. 2015 (in press). Mathematical morphologies based on nondecreasing ε -filters. *Int. J. Comput. Syst. Sci.*

Быстрые алгоритмы обработки изображений на основе нестационарной гамма-нормальной модели

*Грачева Инесса Александровна**

gia1509@mail.ru

Копылов Андрей Валериевич

And.Kopylov@gmail.com

Тула, Тульский государственный университет

Принятый в данной работе байесовский подход часто используется как универсальный метод для целого класса задач обработки изображений. Результат обработки в этом случае может быть представлен как преобразование исходного изображения, определенного на дискретном множестве, во вторичный массив данных, который определен на том же множестве аргументов и принимает значения из некоторого подходящего множества в зависимости от решаемой задачи. Анализируемое изображение и результат обработки в рамках данного подхода рассматриваются как наблюдаемая и скрытая компоненты двухкомпонентного случайного поля. Гамма-нормальная модель скрытого поля, предложенная ранее, хорошо показала себя в задаче удаления шума на изображении. Однако для таких задач, как удаление тумана на изображении, сжатие контраста HDR (high dynamic range) изображений, восстановление структуры изображения, данная постановка не подходит [1].

При решении задач данного типа в качестве наблюдаемой компоненты выступает не исходное изображение, а промежуточный результат обработки, специфичный для каждой конкретной задачи. Например, в задаче удаления тумана в роли наблюдения выступает карта рассеивания, в задаче сжатия контраста — слабоконтрастное изображение, в задаче восстановления структуры — грубая оценка границы. Результирующее изображение получается путем уточнения этого промежуточного изображения на основе процедуры фильтрации-интерполяции с помощью оценок параметров модели скрытого поля, найденных с учетом управляющего изображения, которым может быть исходное изображение.

Работа поддержана грантом РФФИ № 13-07-00529.

- [1] *Грачева И. А. Копылов А. В.* Быстрые алгоритмы обработки изображений на основе нестационарной гамма-нормальной модели // Машинное обучение и анализ данных, 2015. Т. 1. № 12. С. 1677–1685. jmla.org/papers/doc/2015/JMLDA2015no12.pdf.

Fast image processing algorithms on the basis of nonstationary gamma-normal model

*Gracheva Inessa**

author_gia1509@mail.ru

Kopylov Andrey

And.Kopylov@gmail.com

Tula, Tula State University

The Bayesian approach accepted in this paper is often used as a universal framework for a class of image processing problems. The ultimate aim of processing can be represented as a transformation of the original image, defined on a subset of the two-dimensional discrete space, into a secondary array, which would be defined on the same argument set and take values from a set depending on the problem. In this approach, the analyzed image and result of processing are considered as the observed and hidden components of the two-component random field. Proposed earlier gamma-normal model of the hidden field performed well in the image denoising problem. However, this formulation is not suitable for problems such as image haze removal, HDR (high dynamic range) image compression, and structure-transferring of image.

In such problems, the intermediate result of processing, special for each particular problem, plays the role of observation instead of original image. For example, in the haze removal problem, a transmission map is taken as the observation; in the HDR compression problem, observation is a low-contrast image; and in the structure-transferring problem, it is a rough estimate of edge. The resulting image is obtained by refining this intermediate image on the basis of filtration-interpolation procedure using the estimates of parameters of a hidden field. These estimates can be found by taking into account the governor image. In the mentioned above tasks, the original analyzed image can be used as the governor image [1].

This research is funded by the Russian Foundation for Basic Research, grant 13-07-00529.

- [1] Gracheva, I., and Kopylov A. 2015. Fast image processing algorithms on the basis of nonstationary gamma-normal model. *J. Mach. Learn. Data Anal.* 1(12):1677–1685. jmla.org/papers/doc/2015/JMLDA2015no12.pdf.

Нелинейная многомерная фильтрация многокомпонентных изображений

*Медведева Елена Викторовна**

emedv@mail.ru

Трубин Игорь Сергеевич

trubin@vyatsu.ru

Устюжанина Екатерина Александровна

tryp92@mail.ru

Лалетин Алексей Вячеславович

brodilan92@gmail.com

Киров, Вятский государственный университет

Цель работы — разработка метода нелинейной многомерной фильтрации многокомпонентных изображений, позволяющего повысить качество изображений, искаженных белым гауссовским шумом.

Многокомпонентные изображения обладают большой статистической избыточностью, которую целесообразно использовать для повышения качества восстановления зашумленных изображений. Частным случаем многокомпонентных изображений являются цветные RGB изображения, каждая из цветовых компонент которого представляет собой g -разрядное цифровое полутоновое изображение (ЦПИ). Учитывая характер статистической связи между элементами внутри ЦПИ и между элементами цветовых компонент, предложено аппроксимировать цветные изображения трехмерной цепью Маркова с несколькими состояниями, а разрядные двоичные изображения цветовых компонент — трехмерной цепью Маркова с двумя состояниями.

Метод основан на аппроксимации многокомпонентных изображений трехмерной цепью Маркова и теории фильтрации условных марковских процессов. Для повышения качества восстановленного изображения использовано скользящее окно, позволяющее точнее вычислить статистические характеристики изображений для каждой локальной области изображения.

Показана эффективность разработанного метода. В диапазоне отношений сигнал/шум $\rho_{in}^2 = -9 \dots -3$ дБ выигрыш в среднеквадратичном отклонении трехмерным алгоритмом по сравнению с ранее разработанным двумерным алгоритмом составляет от 30% до 70% соответственно [1].

- [1] *Медведева Е. В.* Нелинейная многомерная фильтрация многокомпонентных изображений // Машинное обучение и анализ данных, 2015 (в печати). Т. 1. №.13. jmla.org/papers/doc/2015/JMLDA2015no13.pdf.

Multidimensional nonlinear filtration of multicomponent images

*Medvedeva Elena**

emedv@mail.ru

Trubin Igor

trubin@vyatsu.ru

Ustyuzhanina Ekaterina

tryp92@mail.ru

Laletin Alexey

brodilan92@gmail.com

Kirov, Vyatka State University

The goal of this paper is to develop a method of nonlinear multidimensional multicomponent images filtering. The method allows to improve the quality of image distorted by white Gaussian noise.

Multicomponent images have a greater statistical redundancy. This redundancy would be appropriate for use to improve the quality of the restoration of noisy images. Special cases of multicomponent images are RGB image, each color component of which is a g -bit half-tone digital image (HTDI). The nature of the statistical relationship between elements within the HTDI and among the elements of color components allows to suggest an approximation for the three-dimensional (3D) color images with a Markov chain with several states and for bit binary image of two color components with a 3D Markov chain with two states.

This approximation makes it possible to apply the theory of filtration of conditional Markov processes for the development of filtering method of multicomponent images. The authors propose improving the accuracy of calculation of the statistical characteristics of each local region within the image and between the color components to improve the quality of the reconstructed image. A sliding window is used to estimate local statistical characteristics of the image.

The results of modeling are presented. The gain in the mean square error of a 3D nonlinear filtration with use of the sliding window compared with earlier developed algorithm of a two-dimensional filtration is from 30% to 70%, respectively, in the range of the signal/noise relations, $\rho_{in}^2 = -9 \dots -3$ dB [1].

- [1] Medvedeva, E. 2015 (in press). Multidimensional nonlinear filtration of multicomponent images. *J. Mach. Learn. Data Anal.* 1(13). jmla.org/papers/doc/2015/JMLDA2015no13.pdf.

Использование Радон и Фурье преобразований растровых изображений для описания и отслеживания заданных объектов

Новиков Евгений Анатольевич

eugen@novikov.de

*Падалко Михаил Александрович**

padalkom@gmail.com

Калининград, Институт прикладной математики и информационных технологий БФУ им. И. Канта

Задача мониторинга динамических изображений является одной из основных в области обработки графических данных. Все существующие на сегодняшний день подобные алгоритмы способны решать эту задачу только в очень узких областях, и нет некоего универсального подхода к ее выполнению.

Предлагаемый в данной статье метод позволяет производить анализ изображений и видеоданных по широкому спектру признаков. Он представлен в виде общего описания алгоритма и результатов экспериментальной проверки его эффективности при решении поставленной задачи — быстрой и качественной обработке графических данных в виде динамических изображений или видеопотоков. Рассматриваемый метод реализует новый подход к задаче распознавания образов и основывается на применении комбинации классических методов прямого преобразования Радона к матрице изображения, одномерного преобразования Фурье к полученным интегральным проекциям и статистического анализа интегральных коэффициентов Фурье, рассматриваемых в качестве основных дескрипторов объектов изображения.

Представленный метод отличается универсальностью, т. е. он может применяться для решения задач распознавания любых объектов с любой заранее выбранной глубиной точности. В частности, метод позволяет с высокой точностью отслеживать движения зрачков человека, что находит применение в компьютерной окулографии [1].

- [1] *Новиков Е. А., Падалко М. А.* Использование Радон и Фурье преобразований растровых изображений для описания и отслеживания заданных объектов // *Машинное обучение и анализ данных*, 2015 (в печати). Т. 1. №. 13. jmlida.org/papers/doc/2015/JMLDA2015no13.pdf.

Using Radon and Fourier transformations of raster images for description and tracking of specified objects

Novikov Evgeny

eugen@novikov.de

*Padalko Mikhail**

padalkom@gmail.com

Kaliningrad, Immanuel Kant Baltic Federal University, Institute of Applied Mathematics and Information Technologies

The task of dynamic images monitoring is one of the central ideas in the field of graphic data processing. All currently existing algorithms are able to solve this task only in very narrow fields; there is no general method or approach to its solution.

The proposed method allows analysis of images and video over a wide range of characteristics. This method is presented in the form of a general algorithm description and results of algorithm efficiency testing in addressing the task — fast and quality processing of graphical data as dynamic images or video streams. The proposed method implements a new approach to the task of pattern recognition and is based on a combination of classical methods of direct Radon conversion of image matrix, one-dimensional Fourier conversion to the corresponding projections, and statistical analysis of integral Fourier coefficients considered as objects' features of images.

Another task of data processing, one of the most complex and at the same time pressing problems, is the pattern recognition problem. It belongs to the field of machine vision. The proposed method is versatile, meaning it can be used to solve the problems of recognition of any objects with any preselected depth accuracy. In particular, the method allows highly precise human pupil movement tracking that is used in computer oculoigraphy.

Taking into account the properties of the data obtained by applying the method to a particular image, the conclusions have been drawn about the possible applications of the algorithm and the prospects for its use [1].

- [1] Novikov, E., and Padalko M. 2015 (in press). Using Radon and Fourier transformations of raster images for description and tracking of specified objects. *J. Mach. Learn. Data Anal.* 1(13). jmla.org/papers/doc/2015/JMLDA2015no13.pdf

Комбинированная нелинейная фильтрация цифровых изображений большой разрядности

Петров Евгений Петрович

eppetrov@mail.ru

Харина Наталья Леонидовна

harina@vyatsu.ru

*Ржаникова Елена Дмитриевна**

ran135@mail.ru

Киров, Вятский государственный университет

Требование быстрее передавать большой объем информации, какими являются многоразрядные цифровые изображения (ЦИ), является актуальной задачей, требующей совершенствования средств радиосвязи. Одним из путей сокращения времени передачи ЦИ является переход к многофазным фазомодулированным (ФМ) сигналам, применению которых препятствуют потери помехоустойчивости при каждом делении фазы по сравнению с бинарными ФМ сигналами.

В данной работе для компенсации потерь помехоустойчивости при передаче ЦИ многофазными ФМ сигналами разработан алгоритм восстановления многоразрядных ЦИ, искаженных белым гауссовским шумом (БГШ), за счет эффективной реализации статистической избыточности ЦИ. Например, по сравнению с передачей ЦИ бинарными ФМ сигналами время передачи ЦИ четырехфазными ФМ сигналами сократилось в 2 раза без потерь помехоустойчивости. При передаче ЦИ восьмифазными ФМ сигналами время сокращается в четыре раза, с большой, но не полной компенсацией потерь помехоустойчивости. Используя высокую эффективность реализации статистической избыточности ЦИ разработанного алгоритма, построен комбинированный алгоритм, состоящий из двух алгоритмов: нелинейной фильтрации ЦИ, искаженных БГШ, и медианного фильтра для восстановления ЦИ, искаженных импульсными помехами типа «перец-соль». За счет выделения импульсных помех из БГШ медианный фильтр эффективно подавляет импульсные помехи. Результаты такого сочетания позволяют сократить время передачи многоразрядных ЦИ и успешно бороться с БГШ и импульсными помехами [1].

- [1] *Петров Е. П., Харина Н. Л., Ржаникова Е. Д.* Комбинированная нелинейная фильтрация цифровых изображений большой разрядности // Машинное обучение и анализ данных, 2015 (в печати). Т. 1. №. 13. jmla.org/papers/doc/2015/JMLDA2015no13.pdf

Combined nonlinear filtration of digital high bitness images

Petrov Evgeniy

eppetrov@mail.ru

Kharina Nataliya

harina@vyatsu.ru

*Rzhanikova Elena**

ran135@mail.ru

Kirov, Vyatka State University

The requirement for transfer of a large volume of information, such as a multibit digital images (DI), more quickly is an actual task and demands perfecting of radiocommunication means. One of the ways of a reduction of a DI transfer time is transition to a multiphase frequency modulation (FM) signals. However, their application is limited because of a noise stability loss at each division of a phase in comparison with binary FM signals. At the transfer of DI by the eight-phase FM signals, the time is reduced by four times, but with partial compensation of a noise stability loss.

In this work, the algorithm of restoration of a multibit DI distorted by white Gaussian noise (WGN) is developed. The statistical redundancy of the DI is efficiently used for compensation of a noise stability loss at the transfer of a digital images by multiphase FM signals. For example, the time of the DI transfer by four-phase signals was reduced twice without a noise stability loss in comparison with the DI transfer by the binary FM signals. The combined algorithm of a filtration of multidigit DI is constructed. It consists of two algorithms: a nonlinear filtration of DI distorted by WGN and the median filter for restoration of DI distorted by salt-pepper impulse noise. Due to separation of impulse noise and WGN, the impulse noise is efficiently suppressed by the median filter. The results of such combination allow to reduce transfer time of a multibit DI and to strive successfully against WGN and impulse noise [1].

- [1] Petrov, E. 2015 (in press). Combined nonlinear filtration of digital high bitness images. *J. Mach. Learn. Data Anal.* 1(13). jmla.org/papers/doc/2015/JMLDA2015no13.pdf.

Метод быстрого сжатия изображений без потерь

Петров Евгений Петрович

eppetrov@mail.ru

Харина Наталья Леонидовна

natal_res@mail.ru

Суших Павел Николаевич

sukhikh.pavel@gmail.com

Киров, Вятский государственный университет

В настоящее время использование цифровых изображений (ЦИ) в различных сферах деятельности, таких как медицина, геодезия, картография, системы мониторинга и т. п., привело к необходимости использования различных хорошо известных методов сжатия ЦИ. Однако существует ряд приложений, где использование этих методов не всегда возможно из-за ограничений на вычислительные и энергетические ресурсы, например системы дистанционного зондирования Земли. Большие космические аппараты обладают энергетическими возможностями для обеспечения непрерывного режима передачи информации, требующего больших вычислительных ресурсов на борту. Появившиеся в последнее время малые спутники не могут обеспечить непрерывный режим передачи информации из-за жестких ограниченных требований к эффективности использования энергетических ресурсов, что вызвало необходимость разработки новых энергетически малозатратных методов сжатия ЦИ, не уступающих известным методам сжатия многоразрядных ЦИ [1].

Предложен новый метод сжатия ЦИ, удовлетворяющий перечисленным требованиям. Алгоритм состоит из следующих процедур: разделение ЦИ на разрядные двоичные изображения (РДИ); предсказание каждого элемента РДИ на основе теории условных марковских процессов; кодирование любым известным способом. Полученные результаты исследования эффективности предложенного метода сжатия без потерь показали, что он не уступает аналогам (PNG и JPEG-LS) по быстродействию, обладает большей степенью сжатия и имеет ряд преимуществ:

- возможность параллельной обработки РДИ;
- возможность обработки ЦИ любой разрядности;
- отсутствие вычислительных операций на этапе предсказания.

- [1] *Петров Е. П., Харина Н. Л., Суших П. Н.* Метод быстрого сжатия изображений без потерь // Машинное обучение и анализ данных, 2015 (в печати). Т. 1. №. 12. jmla.org/papers/doc/2015/JMLDA2015no12.pdf.

Fast lossless image compression method

Petrov Eugeny

eppetrov@mail.ru

Kharina Natalya

natal_res@mail.ru

*Sukhikh Pavel**

sukhikh.pavel@gmail.com

Kirov, Vyatka State University

At present, the use of digital images (DI) in various fields of activity, such as medicine, geodesy, cartography, monitoring systems, etc., led to necessity of various digital image compression methods. However, there are some applications where these methods cannot be used because of the computing and energy resources limitation, for example, Earth remote sensing systems that form immense volume of digital data on the board of a space vehicle. Huge space vehicle has energy opportunities to provide a continuous data transmission that require ample resources on the board. The latest miniature space watch facilities (small satellites) cannot provide a continuous data transmission because of the severely constrained requirements for energy resources usage efficiency that produces the necessity of new energy efficient low-cost DI compression methods, which would not be inferior to the known multidigit DI compression methods with high resolution but surpass them [1].

Here, the new DI compression method, which meets the listed requirements, is proposed. The algorithm consists of the following procedures, splitting the DI into binary images (BI), predicting of each element of BI being based on the theory of the conditional Markov processes with discrete states and coding using any known algorithm. The obtained research results indicate that the proposed method performance is not worse than of the known algorithms, and that its compression ratio is higher and it has the following advantages:

- the capability of simultaneous BI processing;
- the capability of DI processing with digit capacity; and
- the lack of computational operations at the stage of prediction.

- [1] Petrov, E. P., Kharina N. L., and Sukhikh P. N. 2015 (in press). Fast lossless image compression method. *J. Mach. Learn. Data Anal.* 1(12). jmla.org/papers/doc/2015/JMLDA2015no12.pdf.

Морфологическая фильтрация изображений на основе взаимного контрастирования

*Рубис Алексей Юрьевич**

arcelt@mail.ru

Лебедев Максим Алексеевич

MLebedev@gosniias.ru

Визильтер Юрий Валентинович

viz@gosniias.ru

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

Для решения задачи сравнения изображений предложены морфологические *фильтры на основе взаимного контрастирования* (ВК-фильтры) вида

$$\psi_a^w(f, g) = g_0^{w(x,y)}(x, y) + a(f, g^{w(x,y)})(g(x, y) - g_0^{w(x,y)}(x, y)),$$

где $a(f, g^{w(x,y)})$ — локальный коэффициент взаимного сходства фрагмента изображения $g^{w(x,y)}$ с элементами изображения f ; $o(x, y) = const$ — простейшее постоянное изображение. В качестве вариантов реализации $a(f, g^{w(x,y)})$ могут применяться: нормированный линейный коэффициент корреляции, коэффициент геометрической корреляции форм, «тепловые» ядра взаимного сходства фрагментов изображения на основе сравнения векторов локальных признаков и др.

Для практической реализации разработан алгоритм нормализации фона на основе взаимного контрастирования с использованием пирамиды изображений. Алгоритм успешно протестирован на реальных изображениях различных типов.

Достоинствами предложенных ВК-фильтров по сравнению с морфологическими проекторами Пытьева является то, что для их работы не требуется предварительная сегментация изображения, а также то, что изменения на изображениях выделяются при существенном изменении как яркости, так и геометрии сцены [1].

Работа поддержана грантом РФФИ № 14-07-00914-а.

- [1] *Рубис А. Ю., Лебедев М. А., Визильтер Ю. В.* Морфологическая фильтрация изображений на основе взаимного контрастирования // Вестник компьютерных и информационных технологий, 2015 (в печати).

Morphological image filtration based on mutual contrast

*Rubis Aleksey**

arcelt@mail.ru

Lebedev Maksim

MLebedev@gosniias.ru

Vizilter Yuri

viz@gosniias.ru

Moscow, The Federal State Unitary Enterprise “State Research Institute of Aviation Systems” (FGUP “GosNIIAS”)

Regarding the task of image comparison, the paper proposes morphological *filters based on mutual contrast* (MC-filters) and described as

$$\psi_a^w(f, g) = g_0^{w(x,y)}(x, y) + a(f, g^{w(x,y)})(g(x, y) - g_0^{w(x,y)}(x, y))$$

where $a(f, g^{w(x,y)})$ is the local mutual similarity coefficient of image fragment $g^{w(x,y)}$ and elements of another image f and $o(x, y) = const$ is the simplest constant image. There are other possible implementations of $a(f, g^{w(x,y)})$ such as linear correlation coefficient, geometry correlation coefficients, image similarity measures based on heat-kernels and diffusion maps, etc.

Related to the practical implementation, the background normalization algorithm based on multiscale mutual contrasting was developed. The algorithm was successfully tested on different real images.

The main advantage of the proposed MC-filters compared to morphological Pyt’ev projector is that their operation does not require the prior segmentation of the image and that changes in the images are distinguished with a significant change in brightness and the scene geometry [1].

This research is funded by the Russian Foundation for Basic Research, grant 14-07-00914-a.

- [1] Rubis, A., Lebedev M., and Vizilter Yu. 2015 (in press). Morphological image filtration based on mutual contrast. *Herald of Computer and Information Technologies*.

Метод повышения разрешения космических изображений с использованием априорной информации в векторной форме для сохранения границ

Трекин Алексей Николаевич^{1*} alexey.trekin@gmail.com

*Матвеев Иван Алексеевич*² matveev@ccas.ru

*Мурынин Александр Борисович*² amurnin@bk.ru

*Бочкарева Валерия Георгиевна*¹

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Вычислительный центр им. А. А. Дородницына РАН

Разработан метод повышения разрешения изображений, использующий векторную информацию о геометрических свойствах объектов, содержащихся на изображениях. Метод сочетает разные подходы к повышению разрешения вблизи известных контрастных границ и вдали от них, что позволяет комбинировать эффективные алгоритмы для оптимального повышения разрешения с сохранением четких, резких границ там, где это необходимо, и гладких, непрерывных решений в прочих местах. Для повышения разрешения вблизи границ использован метод Шепарда интерполяции по нерегулярной сетке, а в удалении от них — метод бикубической интерполяции. Качество повышения разрешения исследовано на космических изображениях Landsat с использованием векторной информации в виде карты водоемов. Результаты численного эксперимента показывают, что использование векторной информации позволяет улучшить результат по сравнению с бикубической интерполяцией. Приведенные результаты позволяют применимость разработанного метода для повышения пространственного разрешения данных дистанционного зондирования Земли [1].

Работа поддержана РФФИ, проекты № 14-01-00348 и № 13-05-12019.

- [1] *Трекин А. Н., Матвеев И. А., Мурынин А. Б., Бочкарева В. Г.* Метод повышения разрешения космических изображений с использованием априорной информации в векторной форме для сохранения границ // *Машинное обучение и анализ данных*, 2015 (в печати). Т. 1. №. 13. jmla.org/papers/doc/2015/JMLDA2015no13.pdf.

A method for upsampling of remote sensing images using vector data for preserving edges

Trekin Alexey^{1*}

alexey.trekin@gmail.com

*Matveev Ivan*²

matveev@ccas.ru

*Murynin Alexander*²

amurynin@bk.ru

*Bochkareva Valerija*¹

¹Moscow, Moscow Institute of Physics and Technology (State University)

²Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

A new method for upsampling of aerial and space images was developed. The method uses vector data about geometry of objects contained in image. *A priori* information about high-contrast boundaries helps to preserve sharp illuminance change from blurring due to up-sampling procedure. The developed method was tested over a set of remote sensing images and vector map of water bodies.

The method uses different approaches to upsampling close to known boundaries and far from them that allows to combine efficient algorithms for optimal upsampling preserving sharp edges where it is necessary while making smooth picture in other regions. In the vicinity of boundaries, the Shepard interpolation technique is used, while bicubic interpolation is a method for upsampling of points far from boundaries. Upsampling quality is assessed using Landsat remote sensing imagery and map of water bodies as vector data. Two measures are used to test upsampling quality: mean square difference and index of structural similarity, both showing improvement compared to bicubic interpolation. The results achieved show that the developed method can be used for upsampling of remote sensing images [1].

This research is funded by the Russian Foundation for Basic Research, grants 14-01-00348 and 13-05-12019.

- [1] Trekin, A. N., Matveev I. A., Murynin A. B., and Bochkareva V. G. 2015 (in press). A method for upsampling of remote sensing images using vector data for preserving edges. *J. Mach. Learn. Data Anal.* 1(13). jmla.org/papers/doc/2015/JMLDA2015no13.pdf.

Минимизация признакового пространства распознавания трехмерного изображения на основе стохастической геометрии и функционального анализа

*Федотов Николай Гаврилович*¹

fedotov@pnzgu.ru

Сёмов Алексей Александрович^{1*}

mathematik_aleksey@mail.ru

*Моисеев Александр Владимирович*²

moigus@mail.ru

¹Пенза, Пензенский государственный университет

²Пенза, Пензенский государственный технологический университет

Предложен новый подход к распознаванию трехмерных изображений, основанный на современных методах стохастической геометрии и функционального анализа. Данный метод обладает рядом преимуществ, в частности позволяет описывать метрические свойства трехмерных объектов. Благодаря построению строгой математической модели, аналитик может строить признаки аналитически, описывая форму объектов и их особенности (например, конструирование геометрических признаков).

Гипертрейс-преобразование позволяет создавать инвариантное описание пространственного объекта, которое является более устойчивым к искажениям и координатным шумам, чем описание, получаемое в результате процедуры нормализации объекта. Достоверность и эффективность предлагаемого метода подтверждается как адекватно построенной математической моделью с применением современных подходов анализа и распознавания трехмерных изображений, так и результатами практических экспериментов.

Дано описание техники сканирования гипертрейс-преобразования и его математической модели. Предложена собственная методика минимизации признакового пространства и соответствующая ей решающая процедура. Приведены результаты практического эксперимента сравнения стохастического и детерминированного способов сканирования [1].

Работа поддержана грантом РФФИ № 15-07-04484.

- [1] *Федотов Н. Г., Сёмов А. А., Моисеев А. В.* Минимизация признакового пространства распознавания трехмерного изображения на основе стохастической геометрии и функционального анализа // Машинное обучение и анализ данных, 2015 (в печати). Т. 1. №. 13. jmla.org/papers/doc/2015/JMLDA2015no13.pdf.

Feature space minimization of three-dimensional image recognition based on stochastic geometry and functional analysis

*Fedotov Nikolay*¹

fedotov@pnzgu.ru

Syemov Aleksey^{1*}

mathematik_aleksey@mail.ru

*Moiseev Alexandr*²

moigus@mail.ru

¹Penza, Penza State University

²Penza, Penza State Technological University

A new approach to the three-dimensional (3D) objects' recognition based on modern methods of stochastic geometry and functional analysis is suggested. This method has many advantages; in particular, it allows to describe 3D objects metric properties. Thus, due to building a rigorous mathematical model, the analyst can construct analytical and not intuitive features, describing object form and their characteristics (in particular, constructing geometric features).

Hypertrace transform allows to create invariant description of spatial object which is more resistant to distortion and coordinate noise than the description obtained as a result of the object normalization procedure. The proposed method reliability and efficiency are confirmed both an adequate constructed mathematical model by using modern approaches of 3D images analysis and recognition and the practical experiments results and, also, the developed software package registration.

In the article, detailed description of hypertrace transform scan technique and its mathematical model is provided. The main approaches to construct and distinguish informative features are analyzed. Own method to minimize the feature space and its appropriate decision procedure are proposed. The practical experiment results of comparing for stochastic and deterministic scan methods are presented [1].

This research is funded by the Russian Foundation for Basic Research, grant 15-07-04484.

- [1] Fedotov, N. G., Syemov A. A., and Moiseev A. V. 2015 (in press). Feature space minimization of three-dimensional image recognition based on stochastic geometry and functional analysis. *J. Mach. Learn. Data Anal.* 1(13). jmla.org/papers/doc/2015/JMLDA2015no13.pdf.

Верификация модели квазиоптимальной сегментации на примере составного изображения

*Харинов Михаил Вячеславович**

khar@iias.spb.su

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

В модели квазиоптимальной сегментации [1] решается задача получения иерархической последовательности разбиений множества пикселей на каждое число кластеров от 1 до N , где N — число пикселей в изображении. *Квазиоптимальные* иерархические приближения ограничены по числу сегментов, близки к изображению по суммарной квадратичной ошибке E и, в зависимости от числа кластеров, описываются выпуклой последовательностью значений E . Так как последовательность оптимальных приближений не иерархическая, их аппроксимация квазиоптимальными приближениями не однозначна. Управляющим параметром в модели является число суперпикселей.

Выходная последовательность приближений получается в три этапа. На первом этапе попарным слиянием смежных сегментов в модели Мамфорда–Шаха вычисляется приближение изображения с заданным числом сегментов, для которых поддерживается операция деления надвое. Затем при заданном числе сегментов минимизируется ошибка E и формируются суперпиксели за счет перераспределения пикселей между сегментами, а также итеративного разделения сегмента надвое в одном месте изображения при слиянии сегментов в другом. На третьем этапе выполняется иерархическая кластеризация суперпикселей методом Уорда.

Верификация модели проводится на стереоснимках сцены, снятых в различном ракурсе и скомпонованных в единое изображение. Минимизация E обеспечивает зрительное улучшение сегментации и единообразия выделения объектов.

Метод минимизации E и способ верификации применимы для улучшения сегментации по различным алгоритмам.

- [1] *Харинов М. В.* Улучшение качества приближения цифрового изображения на основе иерархической сегментации // Вестник Бурятского гос. ун-та. — Улан-Удэ: Изд-во БГУ, 2014. С. 54–57.

Verification of quasi-optimal segmentation model using composite images

*Kharinov Mikhail**

khar@iiias.spb.su

Saint-Petersburg, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences

In the model of quasi-optimal segmentation [1], obtaining of a hierarchical sequence of partitions of N image pixels into each number of clusters from 1 to N is treated. *Quasi-optimal* hierarchical approximations consist of a limited segment number, closely simulate an image in the total squared error E , and correspond to a convex sequence of E values depending on the number of clusters. Since the sequence of optimal approximations in general is nonhierarchical, then the sequence of quasi-optimal approximations is not uniquely determined. To tune the target approximations, the number of *superpixels* in the form of connected segments is used as a control parameter.

The output sequence of approximations is obtained in three stages. In the first stage, the initial image approximation is calculated by pairwise merging of adjacent segments, as in Mumford–Shah model. This approximation contains the specified number of segments reversibly generated as divisible into two. Then, for a given number of segments, the error E minimization and formation of the superpixels are carried out. The superpixels are obtained by the redistribution of pixels between the segments and also by iterative division of segment into two ones in one place of the image when merging the pair of adjacent segments in the other place. In the third stage, the hierarchical clustering of superpixels by Ward's method is performed.

The model is verified by examples of the scene snapshots obtained from different angles and then united into a single image. Minimization of E provides a visual segmentation improvement along with consistent object detection. Combined globally-local segmentation improvement and the technique of segmentation verification are applicable to improve the results of various segmentation algorithms.

- [1] Kharinov, M. 2014. Improving the quality of digital image approximation based on hierarchical segmentation. *Bulletin of Buryat State University*. Ulan-Ude: BSU. 54–57.

Оценивание сложности изображений двумерными вариациями

Чочиа Павел Антонович*

chochial@iitp.ru

Москва, Институт проблем передачи информации РАН

В реальных задачах часто требуется оценка изображения. Если неискаженный сигнал недоступен, критерием может являться оценка сложности результата. Для сигналов, функций и потока сообщений сложность определяется по-разному, но используемые при этом оценки не применимы к изображению (например, они не являются метрическими и не зависят от амплитуды). Для изображения сложность желательно интерпретировать как оценку, отражающую число, размеры и контраст имеющихся деталей. С увеличением числа и контраста деталей сложность должна возрастать, а с увеличением их размеров — уменьшаться.

Распространенной характеристикой сложности одномерной функции является вариация. Известны разные ее обобщения на функции многих переменных. Почти все они сводятся к нахождению одного функционала, в основе которого так или иначе лежит значение модуля градиента функции в точке, поэтому величины и поведение таких многомерных вариаций близки между собой. Их изучение привело к выводу, что функция многих переменных должна характеризоваться несколькими независимыми функционалами. Для двумерной функции такие функционалы были предложены А. С. Кронродом на основе понятия множеств уровня.

Редукцией к дискретным функциям и модификацией функционалов удастся построить две независимые характеристики — показатели *числа* и *размеров* объектов изображения [1]. Проведен анализ оценок при увеличении шума и сглаживании. Предложенная комбинация характеристик отражает структуру изображения и может служить оценкой его сложности. Рассматривается приведение нескольких вариаций к обобщенному функционалу для использования в оптимизационных схемах.

Работа поддержана грантом Российского научного фонда, проект № 14-50-00150.

- [1] Чочиа П. А. Двумерные вариации как средство оценивания сложности изображений // Машинное обучение и анализ данных, 2015. Т. 1. № 12. С. 1660–1676. jmla.org/papers/doc/2015/JMLDA2015no12.pdf.

Image complexity assessment by two-dimensional variations

*Chochia Pavel**

chochia@iitp.ru

Moscow, Institute for Information Transmission Problems
of the Russian Academy of Sciences

The questions of *image complexity* assessment by using two-dimensional (2D) variations are studied. The image complexity is interpreted as an attribute which is specified by the quantity, the sizes, and the visibility of image details. It is desirable that the image complexity to be increased with the growing of the number and the contrast of image details and to be reduced with the growth of sizes of details.

Different methods were proposed for generalizing one-dimensional variation to estimate functions of several variables. Almost all of them result in unique functional which is commonly constructed on the base of the modulus of function gradient in a point. Therefore, their behaviors are similar. It was asserted by A. S. Kronrod that a function of two variables must be characterized by two independent functionals, and corresponding method was put forward by him.

With the reduction of proposed method to digital images, the modified functional which is adequate to image contents is constructed [1]. The image complexity assessment is proposed which consists of two independent terms: the *component quantity index* and the *component size index*.

The change of variations under different image transformations is analyzed. Theoretical conclusions have been confirmed by the experimental explorations. The proposed combination of 2D variations was demonstrated to reflect morphological structure of an image and to assess its complexity.

This research is funded by the Russian Science Foundation, grant 14-50-00150.

- [1] Chochia, P. A. 2015. Two-dimensional variation as an image complexity assessment. *J. Mach. Learn. Data Anal.* 1(12):1660–1676. jmla.org/papers/doc/2015/JMLDA2015no12.pdf

Сегментация и обнаружение контрастных объектов и их применение в навигации роботов

*Кий Константин Иванович**

`konst.i.kiy@gmail.com`

Москва, Институт прикладной математики им. М. В. Келдыша РАН

Дано описание новой техники для сегментации и описания изображений, разработанной автором. Для решения практических задач, возникающих в робототехнике, необходимы средства для автоматического понимания изображений. Будет недостаточно правильно разбить множество пикселей цветного изображения на компоненты, соответствующие частям реальных объектов на нем. Нужно хорошо описать отношение соседства для выделенных компонент, форму их границ и цветовые (яркостные) параметры. Для этих целей для каждого цветного изображения строится структурный граф цветовых сгустков STG и двудольный граф LRG левых и правых контрастных кривых или ростков контрастных объектов на этом структурном графе (LRG — left and right germs of contrast objects).

Разработаны статистические методы оценки формы границы ростков контрастных объектов. Каждый росток есть цепочка окрашенных интервалов в соседних полосах. Форма ростка описывается дискретной цепочкой концов его интервалов. Для ее оценки в работе использованы гистограммные методы выявления закономерностей в направлениях отрезков, связывающих точки концов интервалов данного ростка. Поскольку росток может содержать небольшое число точек и допускается возможность ошибок сегментации, разработанные методы являются более предпочтительными чем аналитические методы типа метода наименьших квадратов.

В работе приводятся приложения развитой техники к анализу естественных ориентиров, возникающих при движении робота в помещениях. Также описаны применения данной техники для автоматического анализа дорожных сцен [1].

Работа поддержана грантами РФФИ № 13-08-01118 и 13-07-00988.

- [1] *Kiy K. I.* Segmentation and detection of contrast objects and their application in robot navigation // Pattern Recognition Image Analysis, 2015. No. 2. P. 338–346.

Segmentation and detection of contrast objects and their application in robot navigation

*Kiy Konstantin**

konst.i.kiy@gmail.com

Moscow, Keldysh Institute of Applied Mathematics of the Russian Academy of Sciences

A new technique for image segmentation and description, developed by the author, is described. To solve practical problems arising in robotics, the facilities for automatic image understanding are required. It is not sufficient to divide correctly the set of pixels of a color image into the components corresponding to the parts of real objects in it. It is required to describe appropriately the adjacency relations for the selected components, the shape of their boundaries, and their color and intensity parameters. For these purposes, for each color image, the structural graph of color bunches STG and the bipartite graph LRG of left and right contrast curves or germs of contrast objects are constructed.

Methods for evaluating the shape of germs of a contrast object are developed. Each germ is a chain of colored intervals in adjacent strips. Its shape is described by a discrete sequence of ends of its intervals. To evaluate the shape of germs, histogram methods for revealing regularities in directions of segments, connecting the points of ends of intervals of a given germ, are developed. Since the germ may contain a small number of intervals and it is possible that there are segmentation errors, the developed methods are more preferable than analytical methods, like the least-squares method.

In this paper, applications of the developed technique to the analysis of natural landmarks arising in indoor robot motion are presented. Applications of this technique to automatic analysis of road scenes are also described [1].

This research is funded by the Russian Foundation for Basic Research, grants 13-08-01118 and 13-07-00988.

- [1] Kiy, K. 2015. Segmentation and detection of contrast objects and their application in robot navigation. *Pattern Recognition and Image Anal.* 2:338–346.

Выделение отличий на мозаичных изображениях на основе референтных фильтров

*Лебедев Максим Алексеевич** MLebedev@gosniias.ru

Рубис Алексей Юрьевич arcelt@mail.ru

Горбацевич Владимир Сергеевич gvs@gosniias.ru

Визильтер Юрий Валентинович viz@gosniias.ru

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

Статья посвящена обоснованию и развитию новой техники построения компаративных диффузных морфологий. Рассматриваются две альтернативные трактовки морфологического проектора Пытьева, ведущие к двум разным способам понимания формы — «абсолютному» и «относительному». Вводится диффузная морфология на основе суперпозиции двух пытьевских проекторов на мозаичные формы. В качестве обобщения данного приема предложена диффузная морфология на основе референтных фильтров.

Для решения задачи выделения отличий в сцене наблюдения на изображении g по отношению к изображению f в случае небольших сдвигов камеры предлагается использовать в схеме нормализации фона вместо проектора P_F *референтный диффузный фильтр на основе суперпозиции двух морфологических проекторов* $P_{GF} = P_G P_F$. Для случаев с более существенным изменением геометрии сцены предлагается использовать вместо референтного фильтра на основе суперпозиции проекторов *референтный фильтр на основе суперпозиции взаимных диффузных мозаичных фильтров*, который при соответствующем выборе признаков может обеспечить определенную устойчивость процедуры выделения изменений на нормализованном изображении к изменению не только яркости, но и геометрии изображений, что показано на примере обработки модельных изображений [1].

Работа поддержана грантом РФФИ № 14-07-00914-а.

- [1] *Лебедев М. А., Рубис А. Ю., Горбацевич В. С., Визильтер Ю. В.* Выделение отличий на мозаичных изображениях на основе референтных фильтров // Вестник компьютерных и информационных технологий, 2015 (в печати).

Detecting differences on mosaic images based on reference filters

*Lebedev Maksim**

MLebedev@gosniias.ru

Rubis Aleksey

arcelt@mail.ru

Gorbatsevich Vladimir

gvs@gosniias.ru

Vizilter Yuri

viz@gosniias.ru

Moscow, The Federal State Unitary Enterprise “State Research Institute of Aviation Systems” (FGUP “GosNIIAS”)

The article is devoted to verification and development of new techniques for constructing comparative diffuse morphologies. Two alternative interpretations of morphological Pyt’ev projector, leading to two different ways of understanding the form — “absolute” and “relative,” have been considered. A diffuse morphology based on the superposition of the two Pyt’ev projectors on mosaic form has been introduced. As a generalization of this technique, a diffuse morphology is proposed based on the reference filters.

To solve the problem of detecting differences in scene through image g related to image f in case of small camera movement, it is suggested to use *reference diffusion filter based on superposition of two morphological projectors* $P_{GF} = P_G P_F$ instead of projector P_F . In case of more substantial changes in the geometry of the scene, it is proposed to use *reference filter based on superposition of relative diffuse mosaic filters* instead of reference filter based on projectors superposition. The suggested solution can provide robustness of differences detection on a normalized image not only to brightness variances but also to changes in geometry of the image in case of proper feature selection. This is shown on examples of simulated images processing [1].

This research is funded by the Russian Foundation for Basic Research, grant 14-07-00914-a.

- [1] Lebedev, M., Rubis A., Gorbatsevich V., and Vizilter Yu. 2015 (in press). Detecting differences on mosaic images based on reference filters. *Herald of Computer and Information Technologies*.

Обнаружение фрагментов изображений земной поверхности на основе субполосного анализа

*Лихошерстный Алексей Юрьевич** likhoshertny@bsu.edu.ru

Жиляков Евгений Георгиевич zhilyakov@bsu.edu.ru

Белгород, НИУ БелГУ

Рассматривается следующая основная постановка задачи анализа изображений земной поверхности (ИЗП): необходимо выделить объекты одного и того же класса, которые в общем случае могут быть расположены в различных частях изображения. При этом в качестве изначального образа указывается соответствующий фрагмент изображения, тогда как остальные идентичные с ним фрагменты должны быть найдены в автоматическом режиме.

Общий принцип решения задачи заключается в формировании решающей функции, представляющей собой меру близости сравниваемых фрагментов, попадание значений которой в критическую область воспринимается, как основание отвергнуть гипотезу об их идентичности на некотором уровне значимости. Ясно, что в рассматриваемых условиях непосредственно по анализируемому ИЗП могут быть получены только оценки вероятностей ошибок первого и второго родов.

В рамках данной работы используется субполосный анализ, когда свойства фрагментов (объектов) описываются с позиций разбиения области пространственных частот (ПЧ) на некоторые подобласти, непрерывные отрезки трансформант Фурье, в которых используются для формирования меры их идентичности. Такой подход позволяет отразить свойство концентрации энергии фрагмента в наборе подобластей ПЧ, совокупная площадь которых составляет малую долю частотной плоскости. В частности, таким образом оказывается возможным уменьшить влияние малознергетических частотных компонент, например порождаемых широкополосными шумами [1].

Работа выполнена в рамках Государственного задания НИУ БелГУ № 358.

- [1] *Лихошерстный А. Ю.* Обнаружение фрагментов изображений земной поверхности на основе субполосного анализа // Машинное обучение и анализ данных, 2015 (в печати).

The discovery of fragments of images of the Earth based on surface subband analysis

*Likhosherstny Aleksey**
Zhilyakov Evgeniy

likhosherstny@bsu.edu.ru
zhilyakov@bsu.edu.ru

Belgorod, Belgorod State University

The paper deals with the following basic formulation of the problem of image analysis of the Earth surface: necessary to select objects of the same class which generally can be located in different parts of the image. In this case, the corresponding image section is indicated as the original image, while the remaining fragments identical with it are to be found in the automatic mode.

The general principle of solving the problem is to form a decision function which is a measure of proximity of the compared fragments, which values entering the critical region are perceived as a reason to reject hypothesis about their identity at a level of significance. It is clear that under these conditions, directly analyzed images can only be obtained by probability assessment of errors of the first and second kinds.

As part of this work, subband analysis is used when the properties of fragments (objects) are described from the point of partition the field of spatial frequency on some subdomains, continuous segments of Fourier transform which are used to generate a measure of their identity. Such an approach allows to reflect the property of energy concentration of fragment in the set of subdomains drive, which total area is a small fraction of the frequency plane. In particular, thus it is possible to reduce the influence of machine low-energy frequency components, such as generated broadband noise [1].

This research is funded by Belgorod State University, grant 358.

- [1] Likhosherstny, A. 2015 (in press). The discovery of fragments of images of the Earth based on surface subband analysis. *J. Mach. Learn. Data Anal.*

Алгоритмы сегментации изображений сложной структуры

Мурашов Дмитрий Михайлович^{1*} d_murashov@mail.ru
*Березин Алексей Владимирович*² berezin_aleks@mail.ru
*Ганебных Сергей Николаевич*¹ sng@ccas.ru
*Иванова Екатерина Юрьевна*² e_ivanova@mail.ru

¹Москва, Вычислительный центр им. А. А. Дородницына РАН

²Москва, Государственный исторический музей

Одним из видов исследований для датировки произведения станковой масляной живописи является выявление особенностей структуры грунта по изображениям микросрезов красочного слоя. Изображения микросрезов имеют достаточно сложную структуру: несколько слоев грунта и краски, частицы пигментов разных размеров и формы, фрагменты основы, блики, текстурные области и объекты, обусловленные особенностями технологии приготовления препарата. Важной информацией для описания грунта являются геометрические и цветовые характеристики слоев и вкраплений. Для получения такой информации необходима достаточно точная сегментация изображения микросреза.

Предложены автоматизированные алгоритмы сегментации на основе алгоритма кластеризации k -средних и алгоритма суперпиксельной сегментации SLIC (Simple Linear Iterative Clustering) в различных цветовых пространствах. Разработана процедура постобработки для объединения сегментированных областей. Созданные алгоритмы тестировались на изображениях из известных баз тестовых изображений, а также на изображениях реальных образцов красочного слоя картин. Результаты тестов показали, что обеспечивается требуемая точность сегментации. Разработанные алгоритмы могут также применяться для выделения характерных объектов при совмещении изображений, полученных в разных спектральных диапазонах [1].

Работа поддержана грантами РФФИ № 15-07-09324 и № 15-07-07516.

- [1] *Мурашов Д. М., Березин А. В., Иванова Е. Ю.* Сравнение изображений картин по информативным фрагментам // Машинное обучение и анализ данных, 2014. № 8. С. 941–948. jmla.org/papers/doc/2014/JMLDA2014no8.pdf.

Techniques for segmenting images with a complex structure

Murashov Dmitry^{1*}

d_murashov@mail.ru

*Berezin Alexey*²

berezin_aleks@mail.ru

*Ganebnykh Sergey*¹

sng@ccas.ru

*Ivanova Yekaterina*²

e_ivanova@mail.ru

¹Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

²Moscow, State Historical Museum

One of the problems of fine-art paintings dating is detection of peculiarities in ground structure using information acquired from microscopic images of paint layer cross sections. Images of cross sections have a complex structure including several ground and paint layers, pigment particles of various form and size, fragments of support, flares, textural regions, and objects conditioned by the specificity of specimen preparation technology. The ground is characterized by geometrical and color features of layers and particles. For extracting these features, an exact segmentation of cross-section images is required.

In this work, automated segmentation techniques based on k -means clustering and superpixel segmentation SLIC (Simple Linear Iterative Clustering) algorithms in different color spaces are proposed. A postprocessing procedure for segmented regions merging is developed. The proposed techniques were tested on images taken from test image databases and on photographs of cross sections taken from real fine-art paintings. The results of tests show that the developed algorithms provide required precision of segmentation. The techniques can also be used for detecting specific objects for registering images acquired in different spectral bands [1].

This research is funded by the Russian Foundation for Basic Research, grants 15-07-09324 and 15-07-07516.

- [1] Murashov, D., Berezin A., and Ivanova Y. 2014. Comparing images of paintings using informative fragments. *J. Mach. Learn. Data Anal.* 8:941–948. jmla.org/papers/doc/2014/JMLDA2014no8.pdf.

Посегментный анализ однородности в задаче распознавания изображений

Савченко Андрей Владимирович^{1,2*} avsavchenko@hse.ru

¹Нижний Новгород, Нижегородский государственный технический университет им. Р.Е. Алексеева

²Нижний Новгород, Национальный исследовательский университет «Высшая школа экономики»

В [1] исследуется проблема недостаточной точности и вычислительной эффективности систем распознавания изображений при наличии большого числа классов и малого количества эталонов для каждого класса. Согласно предложенной методологии посегментного анализа однородности для поступающего на вход объекта осуществляется его предварительная обработка и сегментация. Для каждого выделенного сегмента выполняется извлечение примитивных признаков, таких как НОГ (histogram of oriented gradients). Рассматривая последовательность признаков как простую выборку, определяется мера близости с посегментным анализом однородности и выравниванием сегментов в некоторой окрестности. Для повышения вычислительной эффективности применяется метод максимально правдоподобного направленного перебора. Если полученное решение оказывается недостаточно надежным (в смысле правила Чоу), степень детализации описания входного объекта повышается, после чего процедура классификации повторяется. Если надежное решение не было найдено, итоговое решение выбирается по принципу максимума апостериорной вероятности. Как показали результаты экспериментов, предложенный подход позволяет на 1%–10% повысить точность по сравнению с традиционными классификаторами. Среднее время классификации оказывается в 2–3 раза ниже по сравнению с методом Pyramid НОГ (PHOG). Таким образом, описанная методология позволяет достичь высокой точности в сочетании с приемлемой производительностью для задач распознавания в условиях малых выборок и большого числа классов.

Работа поддержана грантом РНФ № 14-41-00039.

- [1] *Савченко А.В.* Статистическое распознавание образов на основе посегментного анализа однородности // Машинное обучение и анализ данных, 2015. Т. 1. № 11. С. 1500–1516. jmla.org/papers/doc/2015/no11/Savchenko2015Homogeneity.pdf.

Segment homogeneity testing in the problem of image recognition

Savchenko Andrey^{1,2*}

avsavchenko@hse.ru

¹N. Novgorod, Nizhny Novgorod State Technical University

²N. Novgorod, National Research University “Higher School of Economics”

The focus of the paper is on the small-sample size problem in image classification. If the training set contains small number of models per each class, practically all recognition methods are a special case of k-NN (k-nearest neighbor) classifier. In [1], an approach to design classifiers of complex objects by testing of segment homogeneity has been introduced. This approach is based on the probabilistic model of composite object represented by a sequence of independent identically distributed segments. The asymptotic properties of proposed criterion allow to implement sequential hierarchical classification with approximate search of the nearest neighbor to speed up the decision process. At first, the query object is preprocessed and its segmentation is performed. The primitive features, e.g., the HOGs (histograms of oriented gradients), are extracted for each segment. Next, the homogeneity of each segment of the query object and every model is tested. The maximum likelihood directed enumeration method is used to speed up classification by looking for an approximate nearest neighbor. If the decision is unreliable in terms of the Chow’s rule, the description of the query object is refined by decreasing the size of each segment and the classification is repeated. In case of unreliable decisions at all levels, the final class is chosen by the maximum posterior probability principle. Experimental study in constrained face recognition shows that the proposed approach allows to increase accuracy in 1%–10% in comparison with conventional image recognition techniques. It is 2–3 times faster than the pyramid HOG (PHOG) hierarchical classifier. Thus, described methodology allows to achieve high accuracy with sufficient performance in case of small-sample-size problem and medium-sized number of classes.

This research is funded by the Russian Science Foundation, grant 14-41-00039.

- [1] Savchenko, A. 2015. Statistical pattern recognition based on segment homogeneity testing. *J. Mach. Learn. Data Anal.* 1(11):1500–1516. jmla.org/papers/doc/2015/no11/Savchenko2015Homogeneity.pdf.

Быстрое межкадровое преобразование с использованием локальных бинарных паттернов

Вишняков Борис Ваисович

vishnyakov@gosniias.ru

*Горбачев Владимир Сергеевич**

gvs@gosniias.ru

Сидякин Сергей Владимирович

sersid@gosniias.ru

Егоров Антон Иванович

aegorov@gosniias.ru

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

Предлагается эффективное решение задачи вычисления параметров преобразования пары последовательных кадров последовательности изображений применительно к задачам стабилизации или сшивки изображений. Разработанный алгоритм основан на поиске и сопоставлении особых точек. Найденные пары точек используются для определения матрицы аффинного или проективного преобразования старого кадра в новый. На первом этапе алгоритма предлагается модификация алгоритма разности гауссиан, в котором устойчивые к изменению масштаба особые точки ищутся при различных значениях дисперсии смеси гауссиан. Особыми точками являются локальные максимумы функции отклика (модуль разности двух сверток). Для быстрого расчета свертки используются рекуррентные фильтры. Коэффициенты фильтра подбираются эмпирически таким образом, чтобы минимизировать квадрат разности значений импульсной переходной функции рекуррентного фильтра и аппроксимируемой функции Гаусса. На втором шаге вычисляются локальные бинарные паттерны (LBP — local binary patterns) с различными радиусами. Для сравнения особых точек используется комбинация расстояния Хэмминга между LBP и разности функций отклика детектора особых точек. На последнем шаге определяются параметры аффинного преобразования старого кадра в новый с использованием модифицированного алгоритма RANSAC (RANdom SAMple Consensus). Описанный подход может быть применен в сложных условиях съемки при произвольном движении камеры [1].

Работа поддержана грантом РФФИ № 15-07-09362 А.

- [1] *Vishnyakov B. V., Gorbachevich V. S., Sidiakin S. V.* Fast interframe transformation with local binary patterns // Proceedings of SPIE / Eds. J. Beyeler, F. Puento León. — SPIE, 2015. Vol. 9530. 7 p.

Fast interframe transformation with local binary patterns

Vishnyakov Boris

vishnyakov@gosniias.ru

*Gorbatsevich Vladimir**

gvs@gosniias.ru

Sidyakin Sergey

sersid@gosniias.ru

Egorov Anton

aegorov@gosniias.ru

Moscow, The Federal State Unitary Enterprise “State Research Institute of Aviation Systems” (FGUP “GosNIIAS”)

The solution of the problem of transformation evaluation of the two consecutive frames taken from an image sequence is proposed. The described approach is applicable both to the image stabilization problem in surveillance systems and to the image stitching problem in the field of aerial photography. The main requirement for the image stabilization algorithm in surveillance systems is speed. Linear camera motion model is often unacceptable when perspective distortions arise; so, it is necessary to consider affine or projective transformations for both stabilization and stitching tasks. The proposed algorithm belongs to a group of algorithms based on search and comparison of the feature points. The authors use the found feature point pairs for calculation of an affine or projective transformation matrix of the first image to the second one. A modified version of Difference of Gaussians (DoG) algorithm is suggested to find feature points that are stable to scale changes. Recurrent filters for fast image convolution calculation are applied. Local binary patterns (LBP) descriptors with various radii (to make LBP less sensitive to affine transformations) are calculated. A combination of the Hamming distance and the difference between the DoG feature point responses for the feature points matching is proposed. Then, the affine transformation parameters are determined using modified RANSAC (RANdom Sample Consensus) algorithm. The described approach is computationally efficient (over 150 fps) and can be used in difficult conditions (camera noise and low visibility) and for arbitrary camera movement (jitter, shift, rotation, and scale change) [1].

This research is funded by the Russian Foundation for Basic Research, grant 15-07-09362 A.

- [1] Vishnyakov, B., Gorbatsevich V., and Sidyakin S. 2015. Fast interframe transformation with local binary patterns. *Proceedings of SPIE*. Eds. Eds. J. Beyerer, and F. Puento, León. SPIE. Vol. 9530. 7 p.

Синтез алгоритма сегментации динамически изменяющихся сцен

Ипатов Юрий Аркадьевич

ipatov_ya@list.ru

Парсаев Николай Владимирович

nikolai_parsaev@mail.ru

*Тюкаев Андрей Юрьевич**

TyukaevAY@volgatech.net

Йошкар-Ола, Поволжский государственный технологический университет

Изображения динамически изменяющихся сцен (ДИС) встречаются при анализе изменений объектов, сопровождении целей и обнаружении изменений. Так, под ДИС будем понимать изображения в фиксированной пространственной области, на которой объекты могут изменять положение, форму, текстуру и другие атрибуты.

Цель исследования заключается в создании эффективного подхода для анализа ДИС. Для достижения поставленных показателей точности принятия решений было проведено статистический анализ типовых изображений на примере серии изображений динамики роста растений. Определено оптимальное колориметрическое пространство, обладающее наилучшими интегральными характеристиками классификации фон/объект.

Синтезированный алгоритм сегментации, оптимальный по критерию максимального правдоподобия, сводится к следующим действиям: вычислению проекции цвета текущей точки на выбранную плоскость цветового пространства и вычислению для нее величины отношения правдоподобия; нормировке поля отношений правдоподобия к заданному уровню градации и последующей пороговой обработке.

Для оценки эффективности функционирования программной реализации полученного алгоритма проведена серия экспериментов, где средняя ошибка первого рода 0,02, а второго — 0,1. Созданный алгоритм не уступает по точности принятия решений известным методом сегментации [1].

Работа поддержана грантом Президента РФ МК-7290.2015.9.

- [1] *Ипатов Ю. А.* Локализация объектов динамически изменяющихся сцен цветных изображений на сложном фоне // Вестник ПГТУ, 2015. Т. 27. № 3. С. 15–26.

Synthesis segmentation algorithm for dynamically changing scenes

Ipatov Yury

ipatov_ya@list.ru

Parsaev Nikolay

nikolai_parsaev@mail.ru

*Tyukaev Andrey**

TyukaevAY@volgatech.net

Yoshkar-Ola, Volga State University of Technology

Images of dynamically changing scenes (DCS) can be found for analysis in various fields of scientific and technological activities. Dynamics of objects, target tracking, and detection of changes are some examples of problems faced by the researcher. So, under the DCS, the image in a fixed spatial region in which the objects can change the position, shape, texture, and other attributes is understood.

The main purpose of the study is to provide an effective approach to image analysis for DCS. The task is the synthesis of segmentation algorithm based on supervised learning. To achieve the performance accuracy of decision-making, a statistical analysis of the types of images was made on the example of image series for the dynamics of plant growth. Analysis of color models representation of DCS showed that RGB space has the best integral characteristics classification background/object.

Synthesized optimal segmentation algorithm maximum likelihood confined to the following steps: calculation of color projection for the current point to the selected color space plane and the calculation of the likelihood ratio for it; and normalization of field likelihood ratios for a given level of grayscale for visualization and subsequent thresholding.

To evaluate the efficiency of the software implementation of this algorithm, a series of experiments was fulfilled, where the average false positives were 0.03 and false negatives 0.1. Comparative analysis with known algorithms and synthesized segmentation algorithm for the test-class image showed its effectiveness for precision performance [1].

The work was supported by the grant of President RF MK-7290.2015.9.

- [1] Ipatov, Y. A. 2015. Lokalizatsiya ob'ektov dinamicheski izmenyayushchikhsya stsen tsvetnykh izobrazheniy na slozhnom fone [Localization of objects of dynamically changing scenes of color images on a complex background]. *PGTU Bulletin* 27(3):15–26.

Распознавание эмоций, вызываемых музыкой, на основе анализа оптического потока изображений лица человека

*Князь Владимир Владимирович** vl.kinaz@gosniias.ru

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

Проблеме распознавания эмоций человека на основе анализа видеопоследовательностей в настоящее время уделяется большое внимание. Сложность природы эмоций человека обуславливает трудность их классификации и идентификации. В музыкальной психологии были разработаны разнообразные методы формальной классификации эмоций.

Данная работа посвящена распознаванию человеческих эмоций, вызываемых музыкой, на основе анализа оптического потока изображений человеческого лица.

Предложен метод идентификации эмоций, вызываемых музыкой, на основе оценки вектора эмоционального состояния. Для формального представления человеческих эмоций используется двумерное векторное пространство реакция–степень (P–C), в котором реакция представляет положительное или отрицательное отношение к событию, вызвавшему эмоцию, а степень определяет интенсивность эмоции. Четыре характерные точки лица из стандарта MPEG-4 используются для описания выражения эмоции на лице человека.

Представлен разработанный алгоритм прослеживания характерных точек лица. Точные относительные скорости характерных точек лица оцениваются на основе анализа оптического потока и рассматриваются в качестве выходного вектора эмоции. Алгоритм был протестирован с использованием оригинального программного обеспечения и набора видеопоследовательностей лиц волонтеров [1].

- [1] *Kniaz V., Smirnova Z.* Music-elicited emotion identification using optical flow analysis of human face // Int. Arch. Photogramm. Remote Sens. Spatial Inform. Sci., 2015. Vol. XL-5/W6. P. 27–32. doi:10.5194/isprsarchives-XL-5-W6-27-2015. <http://\linebreakwww.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-5-W6/index.html>.

Music-elicited emotion identification using optical flow analysis of human face

*Kniaz Vladimir**

vl.kinaz@gosniias.ru

Moscow, The Federal State Unitary Enterprise “State Research Institute of Aviation Systems” (FGUP “GosNIIAS”)

Human emotion identification from image sequences is highly demanded nowadays. The highly perceptual nature of human emotions leads to the complexity of their classification and identification. The main question arises from the subjective quality of emotional classification of events that elicit human emotions. A variety of methods for formal classification of emotions were developed in musical psychology. This work is focused on identification of human emotions evoked by musical pieces using human face tracking and optical flow analysis.

A technique for identification of music-elicited emotion is proposed. The technique is based on emotion vector estimation using a number of facial features extracted from an optical flow. Two-dimensional valence–arousal (V–A) emotion vector space is used for formal representation of human emotions. In this space, valence represents the positive or negative quality of an emotion and arousal represents an intensity of an emotion. The original 4 facial features from MPEG-4 Face and Body Animation International Standard are used to describe a facial emotion.

Facial feature tracking algorithm used for facial feature speed and position estimation is presented. Accurate relative speeds of facial features were estimated using optical flow analysis. Obtained relative positions and speeds were used as the output facial emotion vector. The algorithm was implemented in original software. The accuracy of the algorithm was tested using video sequences of volunteers listening to music [1].

- [1] Kniaz, V., and Smirnova Z. 2015. Music-elicited emotion identification using optical flow analysis of human face. *Int. Arch. Photogramm. Remote Sens. Spatial Inform. Sci.* XL-5/W6:27–32. doi:10.5194/isprsarchives-XL-5-W6-27-2015. <http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-5-W6/index.html>.

Методы распознавания и отслеживания меток в системе захвата движения

*Князь Владимир Александрович**

knyaz@gosniias.ru

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

Представлены результаты разработки и применения оригинальной системы трехмерного (3D) захвата движения на основе технического зрения. Система основана на фотограмметрических принципах 3D-измерений и обеспечивает высокую скорость получения изображений, высокую точность 3D-измерений и высокий уровень автоматизации обработки получаемых данных. Для достижения высокой точности 3D-измерений используется оригинальная процедура калибровки, имеющая высокую автоматизацию благодаря применению оригинальных кодированных меток для распознавания и измерения координат опорных точек на изображении. Для фотограмметрической системы захвата движения важными характеристиками, определяющими качество системы, являются автоматизация, надежность и удобство для пользователя. Для решения задач распознавания и прослеживания заданных точек объекта разработаны оригинальные методы распознавания меток и их прослеживания по видеопоследовательности в двухмерном (изображение) и 3D пространстве. Предложена оригинальная методика обнаружения и идентификации меток на основе анализа схожести меток и эпполярной геометрии.

Статья представляет описание конфигурации и технических характеристик разработанной системы захвата движения, методы, используемые для калибровки системы. Приведены алгоритмы, разработанные для распознавания меток и их прослеживания в видеопоследовательности. Представлены результаты использования разработанной системы в различных областях применения [1].

Работа поддержана грантом РФФИ № 15-08-99580а.

- [1] *Knyaz V. A.* Scalable photogrammetric motion capture system "Mosca:" Development and application // Int. Arch. Photogramm. Remote Sens. Spatial Inform. Sci., 2015. Vol. XL-5/W6. P. 43–49. doi:10.5194/isprsarchives-XL-5-W6-43-2015. <http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-5-W6/index.html>.

Methods for recognizing and tracing targets for motion capture system

*Knyaz Vladimir**

knyaz@gosniias.ru

Moscow, The Federal State Unitary Enterprise “State Research Institute of Aviation Systems” (FGUP “GosNIIAS”)

Wide variety of applications (from industrial to entertainment) has a need for reliable and accurate three-dimensional (3D) information about object motion. The great potential for obtaining high accuracy and high degree of automation has vision-based system due to progress in image processing and analysis. Paper presents the results of development, evaluation, and application of original scalable vision-based 3D motion capture system. It is based on photogrammetric techniques of 3D measurements and provides high-speed image acquisition, high accuracy of 3D measurements, and high level of automation of captured data. The original camera calibration and external orientation procedure is used to reach high accuracy of 3D measurements. The calibration procedure is highly automated due to applying original coded targets for identifying and measuring image coordinates of reference points. For a vision-based photogrammetric motion capture system, a reliability and a convenience for user are the key features defining the quality of a system. Some techniques are developed for required object point detection and tracking through video sequences in two-dimensional (2D) image space and 3D object space. The original technique for the targets detection and identification is proposed based on similarity analysis of the targets and epipolar based points correspondence determination. The paper describes the outline and technical characteristics of the developed photogrammetric system and techniques used for the system calibration. The original algorithms developed for object point recognizing and tracking in 2D and 3D spaces are presented. Some results of using of the developed photogrammetric motion capture system in various fields of applications are presented [1].

This research is funded by the Russian Foundation for Basic Research, grant 15-08-99580a.

- [1] Knyaz, V. A. 2015. Scalable photogrammetric motion capture system “Mosca:” Development and application. *Int. Arch. Photogramm. Remote Sens. Spatial Inform. Sci.* XL-5/W6:43–49. doi:10.5194/isprsarchives-XL-5-W6-43-2015. <http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-5-W6/index.html>.

Исследование влияния рассинхронизации входных изображений на качество работы алгоритмов поиска структурных различий

*Корнилов Федор Андреевич**

fakornilov@mail.ru

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

Рассматривается влияние рассинхронизации входных изображений на качество работы алгоритмов поиска структурных различий, под которыми понимаются появившиеся или исчезнувшие объекты местности, для разновременных космических снимков земной поверхности. Работа алгоритмов оценивается на паре изображений, представляющих собой смещенные относительно друг друга копии одного и того же снимка, что позволяет выяснить степень устойчивости методов к геометрическому несоответствию входных данных с исключением влияния структурных различий или случайного шума. Приводится описание нового метода сравнения структур изображений, использование которого совместно с алгоритмами поиска структурных различий позволяет уменьшить количество ложных тревог при рассинхронизации входных изображений. Эксперименты на данных, подготовленных путем взаимного сдвига пары изначально геометрически выровненных космоснимков, показывают, что предложенная модификация алгоритмов подходит для прикладного применения [1].

Работа выполнена при частичной финансовой поддержке государственной темы «Экстремальные задачи и алгоритмы теории приближения функций с приложением к проблемам навигации по геофизическим полям и управления антенными системами» (ГР № 01201367475).

- [1] *Корнилов Ф. А.* Исследование влияния рассинхронизации входных изображений на качество работы алгоритмов поиска структурных различий // Машинное обучение и анализ данных, 2015. Т. 1. № 12. С. 1686–1695. jmla.org/papers/doc/2015/JMLDA2015no12.pdf.

Investigation of the impact of misregistration of input images on the accuracy of change detection

*Kornilov Fedor**

fakornilov@mail.ru

Ekaterinburg, Krasovsky Institute of Mathematics and Mechanics
of the Ural Branch of the Russian Academy of Sciences

The article is devoted to the impact of misregistration input images on accuracy of algorithms structural differences detection for multitemporal Earth's satellite images. Here, structural changes mean appeared or disappeared ground's objects. The results of algorithm work evaluate at pair images prepared with the aid of shift relative to each other copies of the same image that allow to explore degree of stability methods to misregistration input data with excluding influence structural changes or random noise. The usage of new method of comparison image structures with the algorithms of structural differences detection allows to reduce the number of false alarms when input images misregistered. The experiments with real data demonstrated that the proposed modification of algorithms is good enough for applications [1].

The work was partially supported by the state budget theme "Extreme problems and algorithms of approximation theory with application to the problems of navigation on geophysical fields and antenna control systems" (GR No. 1201367475).

- [1] *Kornilov, F. A.* 2015. Investigation of the impact of misregistration of input images on the accuracy of change detection. *J. Mach. Learn. Data Anal.* 1(12):1686–1695. jmla.org/papers/doc/2015/JMLDA2015no12.pdf.

Метод обнаружения движущихся объектов в видеопотоке на основе оценки их границ

Медведева Елена Викторовна

emedv@mail.ru

Карлушин Константин Александрович

konstantin.karlushin@gmail.com

*Курбатова Екатерина Евгеньевна**

kurbatovae@gmail.com

Киров, Вятский государственный университет

Цель работы — разработка метода выделения движущихся объектов в видеопотоке на сравнительно неподвижном фоне на основе оценки их границ, реализация которого достигается меньшими вычислительными ресурсами, чем известными методами.

Метод основан на аппроксимации последовательности цифровых полутоновых изображений (ЦПИ) трехмерной цепью Маркова с несколькими состояниями и представлении ЦПИ набором из g разрядных двоичных изображений (РДИ). Для выделения контуров вычисляется количество информации в каждом элементе РДИ в соответствии с различными сочетаниями элементов окрестности и сравнивается с порогом.

Выигрыш во времени обработки разработанным методом выделения контуров по сравнению с известными методами (Канни, Превитта, Робертса, Собела) составляет 2–5,6 раз.

Для решения задачи выделения объекта интереса по полученным контурным точкам использован плотностный алгоритм кластеризации DBSCAN (density-based spatial clustering of applications with noise).

Выигрыш в среднеквадратичной ошибке точности определения координат для разработанного метода по сравнению с известным разностным методом составляет 1,5–2,5 раз.

Метод может быть применен при обработке данных в реальном масштабе времени. Диапазон варьирования размеров объектов интереса в видеопоследовательности может быть большим, а количество движущихся объектов — априорно неизвестным [1].

[1] *Медведева Е. В., Карлушин К. А., Курбатова Е. Е.* Метод обнаружения движущихся объектов в видеопотоке на основе оценки их границ // Машинное обучение и анализ данных, 2015. Т. 1. № 12. С. 1696–1705. jmla.org/papers/doc/2015/JMLDA2015no12.pdf

Method of detection of moving objects in video stream on the basis of object boundaries estimation

Medvedeva Elena

emedv@mail.ru

Karlushin Konstantin

konstantin.karlushin@gmail.com

*Kurbatova Ekaterina**

kurbatovae@gmail.com

Kirov, Vyatka State University

The purpose of the research is to develop a new method of detection of moving objects in the frame sequence using practically stationary background. The method is based on the estimation of moving object boundaries by calculation of the value of information quantity. It requires less computational resources than the existing well-known methods.

The method uses approximation of the digital halftone images (DHI) sequence by a three-dimensional Markov chain with several states and representation of the DHI by g -digit binary images (DBI). To find contours, the value of information quantity in each DBI element is calculated in accordance with various combinations of neighborhood elements and is compared with a threshold.

The developed method is faster in comparison with the known methods (Canny, Prewitt, Roberts, Sobel) in 2–5,6 times.

To define an object of interest by obtained contour points, DBSCAN (density-based spatial clustering of applications with noise) algorithm is used.

The gain in root mean-squared error of coordinates determination accuracy for developed method in comparison with the known subtraction method is 1.5–2.5.

The developed method requires small computational resources; thus, it can be applied in real-time data processing. Range of variation of object dimensions in video sequence can be wide and the number of moving objects can be priori unknown [1].

- [1] Medvedeva, E. V., Karlushin K. A., and Kurbatova E. E. 2015. Method of detection of moving objects in video stream on the basis of object boundaries estimation. *J. Mach. Learn. Data Anal.* 1(12):1696–1705. jmla.org/papers/doc/2015/JMLDA2015no12.pdf.

Композиции признаков для видеотрекинга при помощи фильтра частиц

Нижбицкий Евгений Алексеевич nizhibitsky@cs.msu.ru

Москва, Московский государственный университет
им. М. В. Ломоносова, Факультет ВМК

Рассмотрены модели правдоподобия, основанные на композиции мер сходства извлекаемых из изображений признаков, которые широко используются для задачи отслеживания объектов на видео при помощи фильтра частиц.

Предложены новые способы оптимального многократного извлечения признаков из различных регионов одного и того же изображения. Оптимизация при этом выполняется за счет построения интегральных изображений, впервые примененных в компьютерном зрении для признаков Хаара в алгоритме Виолы–Джонса, для других исследуемых признаков — цветовых гистограмм, гистограмм направленных градиентов и гистограмм на основе локальных бинарных шаблонов.

На примере датасета Bonn Benchmark on Tracking экспериментально показана возможность эффективного использования композиций групп признаков при неэффективности использования каждой группы в отдельности.

С помощью рассмотренных композиций признаков достигнуто качество трекинга, сравнимое с более сложными по своей структуре методами, основанными на построении ансамблей классификаторов с помощью бустинга, и превышающее результаты схожей работы с использованием метода каскадов.

Работа поддержана грантом РФФИ № 14-07-00965.

- [1] *Нижбицкий Е. А.* Композиции признаков для видеотрекинга при помощи фильтра частиц // Машинное обучение и анализ данных, 2015. Т. 1. № 11. С. 1517–1528. jmla.org/papers/doc/2015/no11/Nizhibitsky2015Composition.pdf.

Feature composition in video tracking using particle filters

*Nizhibitsky Evgeny**

nizhibitsky@cs.msu.ru

Moscow, Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics

This work aims at analysis of probability models based on similarity measures of features extracted from images which are widely used in the field of video tracking using particle filters.

New computationally optimal methods for multiple feature extraction from several regions of the same image are proposed. The optimization is performed by using integral images, first prominently used in computer vision within Viola–Jones object detection framework for Haar rectangles, for other studied features — color histograms, histograms of oriented gradients, and histograms based on local binary patterns.

It is experimentally demonstrated that feature compositions can be used even in the tasks where each of them are useless by itself that was illustrated by Bonn Benchmark on Tracking tests.

The performance achieved using proposed compositions is greater than one in the similar study based on cascade technique and comparable to the performance of more complicated models based on ensemble boosting.

This research is funded by the Russian Foundation for Basic Research, grant 14-07-00965.

- [1] Nizhibitsky, E. 2015. Feature composition in video tracking using particle filters. *J. Mach. Learn. Data Anal.* 1(11):1517–1528. jmla.org/papers/doc/2015/no11/Nizhibitsky2015Composition.pdf.

Оптимизация метода компьютерной окулографии для исследований центральной нервной системы на основе пассивного анализа видеоизображения

Новиков Евгений Анатольевич^{1*} eugen@novikov.de
*Ваколюк Ирина Анатольевна*² i.vakoliuk@iris-check.com

¹Калининград, Институт прикладной математики и информационных технологий БФУ им. И. Канта

²Калининград, Химико-биологический институт БФУ им. И. Канта

Процесс регистрации и последующего анализа произвольных и спровоцированных движений глаз в общем случае называется окулографией. Большинство существующих систем компьютерной окулографии опираются на активное сканирование глаз, что означает, как правило, очень дорогостоящее решение. Предлагаемый метод использования стандартной цифровой оптики может быть применен для точной и надежной локализации центров зрачков с помощью градиентов изображения, полученных с частотой не менее 30 кадр/с. Анализ графического отображения изменения положения центра зрачка во времени — окулограммы — позволяет судить о функциональном состоянии отделов головного мозга, вовлеченных в процесс программирования и регуляции движений глаз, в частности экстрапирамидной системы. В отличие от многих других предлагаемый метод компьютерной окулографии является бесконтактным, что особенно важно при работе с детьми и пациентами с неврологическими и психическими расстройствами, дистантным, доступным и не требует специальной подготовки от медицинского персонала или исследователя.

- [1] *Новиков Е. А., Ваколюк И. А., Ахапкин Р. Д., Варчак И. А., Шалагинова И. Г., Швайко Д. А., Буденкова Е. А.* Оптимизация метода компьютерной окулографии для исследований центральной нервной системы на основе пассивного анализа видеоизображения // *Машинное обучение и анализ данных, 2015 (в печати). Т. 1. №. 12.* jmla.org/papers/doc/2015/JMLDA2015no12.pdf.

Optimization method of computer oculo- graphy for research of the central nervous system based on passive video analysis

*Novikov Evgeny*¹*

eugen@novikov.de

*Vakoliuk Irina*²

i.vakoliuk@iris-check.com

¹Kaliningrad, Immanuel Kant Baltic Federal University, Institute
of Applied Mathematics and Information Technologies

²Kaliningrad, Immanuel Kant Baltic Federal University, Institute
of Chemistry and Biology

Some current investigations in physiology and medicine are focused on human analysis of visual information. The process of recording and subsequent analysis of arbitrary and caused eye movements in general is called oculo-
graphy. Most of the existing computer oculo-
graphy systems rely on active eye scanning, which is usually a very expensive solution. The presented method of using standard digital optics can be employed for accurate and reliable localization of eye centers using vector field of image gradients, obtained with frequency of not less than 30 frame/s. Analysis of visualization of changing pupil center position in time (oculogram) allows one to define functional condition of the brain divisions involved in the programming process and regulation of eye movements, in particular, the extrapyramidal system. Unlike many others, the proposed method of computer oculo-
graphy is noncontact, which is especially important when dealing with children and patients with neurological and psychiatric disorders, distant, affordable, and does not require special training of medical staff or researcher.

- [1] Novikov, E., Vakoliuk I., Akhapkin R., Varchak I., Shalaginova I., Shvaiko D., and Budenkova E. 2015 (in press). Optimization method of computer oculo-
graphy for research of the central nervous system based on passive video analysis. *J. Mach. Learn. Data Anal.* 1(12). jmla.org/papers/doc/2015/JMLDA2015no12.pdf.

Вероятностный подход для отслеживания траекторий людей на видео с камер наблюдения

Цымбал Дмитрий Александрович^{1*}

dmitry.tsymbal@antaressoftware.ru

Чайкин Алексей Витальевич^{1,2}

alexey@asgvisum.com

¹Великий Новгород, ASGVisum

²Нижний Новгород, ASGVisum

Задача обнаружения и отслеживания людей на видео с камер наблюдения чрезвычайно актуальна и имеет множество применений в области безопасности и маркетинга.

Существует много методов решения данной проблемы. Наиболее часто встречаются подходы, основанные на «вычитании фона» с последующим сегментированием отличающихся областей и отслеживанием их положения путем обнаружения геометрических пересечений на соседних кадрах. Среди недостатков данного метода следует выделить необходимость установки отдельных вертикально расположенных камер, необходимость калибровки, чувствительность к изменению условий освещения, облакам, теням, бликам, колышущейся листве деревьев и, что особенно важно, к перекрытию объектов.

В данной работе предлагается альтернативный стандартному алгоритм обнаружения и отслеживания людей, основанный на объединении информации о возможных положениях людей из разных источников (детектор силуэта, выделение областей фона) в двумерные массивы — карты вероятности.

Предложенный метод не требует калибровки, более устойчив к изменениям условий освещения, теням, колышущимся ветвям и листве деревьев и не требует установки отдельных, вертикально направленных камер, позволяя использовать стандартные камеры систем безопасности. Проведенный авторами сравнительный эксперимент, показывает преимущества данного подхода перед стандартным на сценах различной степени загруженности.

Probabilistic approach to people tracking in surveillance video

Tsybal Dmitry^{1*}

dmitry.tsybal@antaressoftware.ru

Chaykin Alexey^{1,2}

alexey@asgvisum.com

¹Veliky Novgorod, ASGVisum

²Nizhny Novgorod, ASGVisum

Stable and accurate people tracking on surveillance videos is one of the most challenging computer vision problems that has many applications in security and marketing.

The majority of existing people trackers use foreground extraction with consequent segmentation and overlap based tracking, which results in sensitivity to lighting conditions changes, shadows, reflections, and — most importantly — to overlapping objects (people standing close to each other, etc.).

The authors present a probabilistic approach to people tracking that uses HOG based cascade of boosted classifiers in addition to foreground extraction. By combining the information from FG and HOG to the maps of probable person locations, it is possible to correct both detection and foreground errors and fill in gaps where the results are missed.

The conducted experiment demonstrates the ability of the proposed method to handle overlapping people and tolerate shadows, reflections, and other lighting conditions changes.

Алгоритм последовательной жадной оптимизации целевой функции кодирования для создания биометрических шаблонов в задаче распознавания лиц

*Горбачев Владимир Сергеевич**

gvs@gosniias.ru

Визильтер Юрий Валентинович

viz@gosniias.ru

Воротников Андрей Валерьевич

andronzord@gmail.com

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

В работе [1] рассматривается проблема построения биометрических шаблонов в задаче распознавания лиц. Под биометрическим шаблоном подразумевается некоторое компактное описание цифрового изображения лица. Предполагается строить биометрические шаблоны как бинарные вектора (так называемые «хеш»), а для их сравнения использовать расстояние Хэмминга, т. е. в общем случае задачу построения таких шаблонов можно представить как задачу хеширования с обучением по выборке. Ключевая идея предлагаемого подхода заключается в том, чтобы последовательно «выращивать» хеш-код, добавляя на каждой итерации к имеющемуся уже k -битовому хешу еще один элементарный хеш таким образом, чтобы целевой критерий кодирования при этом монотонно улучшался. Для обучения и тестирования используется размеченная «учителем» база примеров. В качестве элементарных кодеров (кодирующих один бит) используются линейные классификаторы в пространстве признаков. Для обучения элементарных кодеров используется оригинальная модификация алгоритма RANSAC (random sample consensus). Представленные результаты численных экспериментов показывают достижение высокого качества идентификации лиц при сильных ограничениях на размер биометрического ключа.

- [1] *Горбачев В. С., Воротников А. В., Визильтер Ю. В.* Алгоритм последовательной жадной оптимизации целевой функции кодирования для создания биометрических шаблонов в задаче распознавания лиц // Вестник компьютерных и информационных технологий, 2015 (в печати).

Sequential greedy optimization of coding function for binary templates construction in face recognition

*Gorbatsevich Vladimir**

gvs@gosniias.ru

Vizilter Yuri

viz@gosniias.ru

Vorotnikov Andrew

andronzord@gmail.com

Moscow, The Federal State Unitary Enterprise “State Research Institute of Aviation Systems” (FGUP “GosNIIAS”)

The problem of biometric templates generation for face recognition is discussed in [1]. Biometric template is a compact description of face image. It is assumed that biometric templates would be built like binary vectors (“hashes”) and in order to compare them, a Hamming distance would be used. In other words, the problem of building of such templates can be generally presented as hashing problem with selective training. The main idea of this approach is that it is possible to “grow” a hash-code consequently, adding one 1-bit elementary hash to existing k -bit hash with each new iteration in such a way that coding objective criterion would be developed steadily. Labeled by the “teacher” training set is used for training and testing. Linear classifiers are used in feature space as elementary coders (which code just 1 bit). An original modification of RANSAC (random sample consensus) algorithm is used for elementary coders training. The results of numerical experiments presented in the paper show the achievement of high quality of facial identification in case of great restrictions of the biometrical key size.

- [1] Gorbatsevich, V., Vizilter Yu., and Vorotnikov A. 2015 (in press). Algoritm posledovatel'noy zhadnoy optimizatsii tselevoy funktsii kodirovaniya dlya sozdaniya biometricheskikh shablonov v zadache raspoznavaniya lits [Sequential greedy optimization of coding function for binary templates construction in face recognition]. *Vestnik komp'yuternykh i informatsionnykh tekhnologiy* [Herald of Computer and Information Technologies].

Использование алгоритмов глубокого обучения с учителем для оценки возраста человека по фотографии лица

*Дробных Клим Андреевич**

klim.drobnyh@gmail.com

Половинкин Алексей Николаевич alexey.polovinkin@gmail.com

Нижегородский государственный университет
им. Н. И. Лобачевского

Автоматическая оценка возраста человека по фотографии лица является актуальной задачей в настоящее время. Ее решение поможет улучшить взаимодействие между человеком и компьютером, реализовать контроль доступа, основанный на определении возраста и т. д.

Эта задача является классическим примером распознавания образов. Одним из современных подходов к решению задач такого рода является использование алгоритмов глубокого обучения. Ключевой особенностью этих алгоритмов является обучение признакового описания объектов, которое лучше подходит для решения конкретной задачи. В данной статье предлагается использовать алгоритм глубокого обучения с учителем для улучшения точности существующих подходов для оценки возраста. В качестве базовых подходов были выбраны активные модели внешнего вида и биологически обусловленные признаки, так как основанные на них алгоритмы показали самую высокую точность. Для решения задачи классификации использовался каскад бинарных машин опорных векторов с гауссовым ядром.

В рамках данной работы были реализованы базовые подходы и проведены эксперименты. Для сравнения качества работы алгоритмов использовалась популярная среди исследователей данной темы база лиц FG-NET. Эксперименты показали, что добавление признаков, полученных с помощью глубокого обучения с учителем, к признакам существующих подходов может улучшить точность распознавания. Например, для активных моделей внешнего вида улучшение точности составило 4%.

Using Unsupervised Deep Learning for Human Age Estimation Problem

*Drobnyh Klim**

klim.drobnyh@gmail.com

Polovinkin Alexey

alexey.polovinkin@gmail.com

Nizhny Novgorod, Lobachevsky State University of Nizhny Novgorod

Automatic age estimation from facial images is a challenging problem upcoming in recent years. The solution can improve a human-computer interaction, age-specific access control, etc.

This task is a classic example of pattern recognition task. One of the state-of-the-art approaches in this field is a deep learning. The key idea is learning the best feature representation to solve a certain problem. In this paper, it is proposed to use supervised deep learning features to improve accuracy of existing age estimation algorithms. The active appearance model and the bio-inspired features were chosen to be the basic approaches, because they have shown the best accuracy. A cascade of binary support vector machines with gaussian kernel was used as the classifier.

In this work, two basic approaches were implemented. Popular Face and Gesture Recognition Research Network (FG-NET) database was used to estimate an accuracy. The experiments have shown that adding supervised deep learning features may improve an accuracy for some basic models. For example, adding the features to the active appearance model gave the 4 percent gain.

Поиск внешней и внутренней границ радужной оболочки на изображении глаза методом парных градиентов

Ефимов Юрий Сергеевич^{1*}

yuri.efimov@phystech.edu

*Матвеев Иван Алексеевич*²

matveev@ccas.ru

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Вычислительный центр им. А. А. Дородницына РАН

Для поиска внешней и внутренней границ радужки, представленных как окружности, используется модификация методологии Хафа — метод парных градиентов [1]. После обработки изображения фильтром Кэнни из пикселей получившихся границ выбираются пары, с высокой вероятностью лежащие на одной окружности. Вводятся условия выбора и вероятностные коэффициенты качества пары для сокращения перебора. В преобразовании Хафа используются два аккумулятора: двумерный, изоморфный исходному изображению, голосование в котором производится центрами отрезков, задаваемых парами, и одномерная гистограмма диаметров, в которой откладываются длины этих отрезков. Проведен вычислительный эксперимент, проверяющий работоспособность алгоритма на данных из открытых баз изображений радужки. Определены недостатки алгоритма, вызывающие некорректную обработку некоторых входных изображений. Требуется дальнейший анализ предлагаемого алгоритма и повышение его устойчивости.

Работа поддержана грантом РФФИ № 13-01-00866.

- [1] *Ефимов Ю. С.* Поиск внешней и внутренней границ радужной оболочки на изображении глаза методом парных градиентов // *Машинное обучение и анализ данных*, 2015 (в печати). Т. 1. №. 14. jmla.org/papers/doc/2015/JMLDA2015no14.pdf.

Locating external and internal borders of the iris in eye image by paired gradients method

Efimov Yury^{1*}

yuri.efimov@phystech.edu

*Matveev Ivan*²

matveev@ccas.ru

¹Moscow, Moscow Institute of Physics and Technology (State University)

²Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

To search for circular representations of inner and outer boundaries of the iris, a method of paired gradients is used, which is a modification of the Hough methodology [1]. Image is processed with Canny filter and from the resulting boundaries, pairs of pixels are selected which have high probability to belong to one circle. Selection criteria and probability coefficients are introduced for reduction of number of these pairs. The Hough transform uses two accumulators: the two-dimensional isomorphic to the original image in which voting is done by centers of segments defined by pixel pairs and one-dimensional histogram of the diameters where lengths of these segments are collected. Computational experiment is performed to check the efficiency of the algorithm on data from the public iris image databases. Drawbacks of the algorithm that may cause incorrect handling of some of the input images are identified. Further analysis of the proposed algorithm and increasing its stability are required.

This research is funded by the Russian Foundation for Basic Research, grant 13-01-00866.

- [1] Efimov, Y. 2015 (in press). Locating external and internal borders of the iris in eye image by paired gradients method. *J. Mach. Learn. Data Anal.* 1(14). jmla.org/papers/doc/2015/JMLDA2015no14.pdf.

Комбинирование одноклассовой цветовой сегментации и скелетизации бинарных изображений для детектирования кисти руки

Зенин Дмитрий Геннадьевич^{1*} dzendmitry@gmail.com

*Середин Олег Сергеевич*¹ oseredin@yandex.ru

*Кушнир Олеся Александровна*¹ kushnir-olesya@rambler.ru

*Ларин Александр Олегович*² ekzebox@gmail.com

¹Тула, Тульский государственный университет

²Москва, Московский физико-технический институт (ГУ)

В большом количестве работ, посвященных распознаванию жестов и управлению с их помощью различными объектами, вопрос детектирования кисти руки человека в видеопотоке остается за пределами исследований, поскольку сегментирование кисти является достаточно трудной самостоятельной задачей. Сложность ее решения заключается в том, что руки различаются по форме, а кисть человека имеет большое количество степеней свободы.

Существующие стандартные методы решения, использующие вычитание фона или сегментацию кожи по уровню в цветовом пространстве HSV, хорошо работают в условиях однородного неизменяющегося фона. Однако по мере его усложнения выполнить сегментацию становится проблематично.

В ходе исследований был предложен метод решения задачи сегментации с использованием одноклассового пиксельного классификатора, работающего в цветовом пространстве RGB [1], что позволяет найти области, в которых потенциально может находиться кисть руки. Для определения того, в каких именно областях находится кисть, а какие содержат шум, используется сопоставление с образцом на основе скелетных графов бинарных растровых изображений.

Работа поддержана грантами РФФИ 14-07-00527-а и 14-07-31271-мол_а.

- [1] *Larin A., Seredin O., Kopylov A., Kuo S. Y., Huang S.-C., Chen B. H.* Parametric representation of objects in color space using one-class classifiers // Machine learning and data mining in pattern recognition / Ed. P. Pernier. — Lecture notes in computer science ser. — Switzerland: Springer International Publishing, 2014. Vol. 8556. P. 300–314. http://link.springer.com/chapter/10.1007/978-3-319-08979-9_23.

Combining of one-class color segmentation and binary images skeletonization for hand detecting

Zenin Dmitry^{1*}

dzendmitry@gmail.com

*Seredin Oleg*¹

oseredin@yandex.ru

*Kushnir Olesia*¹

kushnir-olesya@rambler.ru

*Larin Alexander*²

ekzebox@gmail.com

¹Tula, Tula State University

²Moscow, Moscow Institute of Physics and Technology (State University)

There are a lot of works devoted to gesture recognition and control of different objects with their help. However, the task of detecting the human's hand in the video stream remains beyond the research, because segmenting of the palm is a quite difficult independent scientific issue. The complexity of its solution is based on a large number of freedom degrees a human's hand, its shape, and color differences. Additional problems arise from the heterogeneity, varying backgrounds.

Existing standard methods use background subtraction or skin segmentation by the level in the HSV color space. Most of them work well in a homogeneous constant background. However, it becomes difficult to perform its segmentation at the increasing complexity.

During the research, a method of resolving problem of segmentation via one-class pixel classifier that works in RGB color space has been suggested [1].

Segmentation allows to find area in which the hand can be located potentially. To determine which specific areas contain a palm and which contain the noise, pattern matching based on the skeletal graph of binary raster images has been used.

This research is funded by the Russian Foundation for Basic Research, grants 14-07-00527 and 14-07-31271.

- [1] Larin, A., Seredin O., Kopylov A., Kuo S.Y., Huang S.-C., and Chen B.H. 2014. Parametric representation of objects in color space using one-class classifiers. *Machine learning and data mining in pattern recognition*. Ed. P. Perner. Lecture notes in computer science ser. Switzerland: Springer International Publishing. 8556:300–314. http://link.springer.com/chapter/10.1007/978-3-319-08979-9_23.

Анализ результатов оконтуривания левого желудочка сердца на эхографических изображениях у здоровых пациентов с помощью автоматического алгоритма

Зюзин Василий Викторович^{1*} zvvzuzin@gmail.com
*Поршнева Сергей Владимирович*¹ sergey_porshnev@mail.ru
*Бобкова Анастасия Олеговна*¹ iconismo@gmail.com
*Мухтаров Андрей Александрович*¹ andrew443209993@yandex.ru
*Бобков Владимир Валерьевич*² btow@yandex1.ru

¹Екатеринбург, Уральский федеральный университет имени первого Президента России Б. Н. Ельцина

²Верхняя Пышма, Медицинские информационные технологии

Обсуждаются результаты построения границ левого желудочка сердца на эхографическом изображении апикальной четырехкамерной проекции сердца человека без патологий сердечной мышцы с помощью автоматического алгоритма оконтуривания. Работоспособность исследованного алгоритма подтверждена экспертными оценками. Предложен критерий, позволяющий автоматически идентифицировать контуры правильной и неправильной формы. Определены направления дальнейших исследований для улучшения качества оконтуривания [1].

Работа выполнена при финансовой поддержке ФГБУ «Фонд содействия развитию малых форм предприятий в научно-технической сфере» в рамках госконтракта № 11475р/20975.

- [1] *Зюзин В. В., Поршнева С. В., Бобкова А. О., Мухтаров А. А., Бобков В. В.* Анализ результатов оконтуривания левого желудочка сердца на эхографических изображениях у здоровых пациентов с помощью автоматического алгоритма // Машинное обучение и анализ данных, 2015. Т. 1. № 11. С. 1529–1538. jmla.org/papers/doc/2015/no11/Zyuzin2015Contouring.pdf.

The analysis of results of the left ventricle contouring using automatic algorithm on ultrasound images

Zyuzin Vasily^{1*}

zvvzuzin@gmail.com

Porshnev Sergey¹

sergey_porshnev@mail.ru

Bobkova Anastasia¹

iconismo@gmail.com

Mukhtarov Andrey¹

andrew443209993@yandex.ru

Bobkov Vladimir²

btow@yandex1.ru

¹Yekaterinburg, Ural Federal University named after First President of Russia B. N. Yeltsin

²Verhnyaya Pyshma, Medical information technologies

The features of automatic contouring of the left ventricle (LV) on echographic sequences are discussed. The automatic algorithm contouring of the LV of the heart on frames, which contain the image of the apical four-chamber projection of the human heart, is proposed. The quality assessment of automatic contouring is obtained with comparison of expert contours and automatically generated contours. It is shown that using parameters such as precision and recall traditionally used in assessing the quality of information search, when comparing expert and automatically generated contours, cannot obtain, adequate from the physical point of view, assessments of the quality of the automated algorithm. The study results in the kinematics of the mass center of the LV region of the heart, which allowed to propose a criterion for the automatic evaluation of the proper construction of an LV contour on separate frames of the video sequence. Identified areas for further research are aimed at improving the quality of contouring [1].

The work was carried out with financial support of Federal State-Financed Establishment The Fund of Advancement of Small Enterprises in Scientific and Technical Sphere within the framework of state contract 11475p/20975.

- [1] Zyuzin, V., Porshnev S., Bobkova A., Mukhtarov A, and Bobkov V. 2015. The analysis of results of the left ventricle contouring using automatic algorithm on ultrasound images. *J. Mach. Learn. Data Anal.* 1(11):1529–1538. jmla.org/papers/doc/2015/no11/Zyuzin2015Contouring.pdf.

Определение зеркальной симметрии фигур на основе цепочек скелетных примитивов

Кушнир Олеся Александровна kushnir-olesya@rambler.ru
Середин Олег Сергеевич oseredin@yandex.ru

Тула, Тульский государственный университет

Задача определения симметрии и нахождения меры симметрии на цветных и яркостных изображениях хорошо известна и обычно решается на основе определения симметрии особых точек. В то же время аналогичная задача для бинарных изображений не ставится, хотя анализ форм изображений показывает, что ряду объектов присуща зеркальная симметрия. Например, в биологии такая симметрия называется билатеральной, а зеркально-симметричные виды, например, бабочки, — билатералами. Более того, симметрия понимается как показатель здоровья видов. Например, ярко выраженное отсутствие симметрии билатеральных листьев растений является индикатором отклонения условий произрастания от нормы. Следовательно, существуют приложения для определения симметрии форм.

Метод сравнения бинарных изображений на основе цепочек скелетных примитивов, предложенный в [1], применим не только для вычисления меры различия фигур, но и для определения симметрии формы и меры симметричности. Скелет симметричных (или квазисимметричных) форм также симметричен, т. е. состоит из симметричных «правой» и «левой» частей. Если построить для этих частей цепочки примитивов, обойдя левую против часовой стрелки, а правую — по часовой, то получим одинаковые или почти одинаковые левую и правую подцепочки, меру симметричности которых можно определить при помощи ранее предложенной функции сравнения цепочек на основе парного выравнивания. Следовательно, решается проблема определения начальной и конечной вершин обхода правой и левой частей скелета, которые будут точками искомой оси симметрии.

Работа поддержана грантами РФФИ №№ 14-07-31271-мол_а и 14-07-00527-А.

- [1] *Kushnir O., Seredin O.* Shape matching based on skeletonization and alignment of primitive chains // 4th Conference (International) on Analysis of Images, Social Networks, and Tests Proceedings. Springer, 2015 (in press).

Reflectional symmetry of shape based on skeleton chains of primitives

*Kushnir Olesia**

kushnir-olesya@rambler.ru

Seredin Oleg

oseredin@yandex.ru

Tula, Tula State University

The problem of symmetry extraction and symmetry measure calculation for gray-scale and color images is well-known. It is usually solved using information about characteristic points of the image. Nevertheless, the same task for the binary images is not formulated despite the fact that the range of shapes of real-world objects has the property of reflectional symmetry. For example, this property is essential for many plant leaves or butterfly species. Such species are known as Bilateria in biology. Moreover, bilateral symmetry in biology is considered as the feature of healthy state. If the shape of bilateral leaves deviates from the symmetry, it means that the environment conditions differ a lot from the optimum. Above-mentioned facts prove that there is a range of applications for the shape symmetry extraction and measuring techniques.

The method for binary images comparison based on primitive chains proposed in [1] can be applied not only to the shape matching but also to the shape symmetry extraction and symmetry measure calculation. The skeleton of symmetric (or almost symmetric) shape also has symmetric property, that is, it can be divided into symmetric “left” and “right” parts. Two chains of primitives can be found for them by traversing right skeleton part clockwise and left part — anticlockwise. These chains are equal or one is very similar to another, they can be compared using pairwise alignment procedure proposed in [1]. Thus, the start and the end skeleton vertices for the “left” and “right” traversing have to be evaluated. These vertices are the points on the symmetry axis, and the measure of “left” and “right” chain parts matching is considered as the symmetry measure of the shape.

This research is funded by the Russian Foundation for Basic Research, grants 14-07-31271 and 14-07-00527.

- [1] *Kushnir O., Seredin O.* 2015 (in press). Shape matching based on skeletonization and alignment of primitive chains. *4th Conference (International) on Analysis of Images, Social Networks, and Tests Proceedings*. Springer.

Нормализация изображений планарии методом жирных линий в биологических исследованиях

Местецкий Леонид Моисеевич^{1*}

mestlm@mail.ru

*Тирас Харлампий Пантелеевич*²

tiras1950@yandex.ru

¹Москва, Московский государственный университет
им. М.В. Ломоносова, Факультет ВМК

²Пушино, Пушкинский естественно-научный институт

Ресничные плоские черви планарии обладают выдающимися способностями к регенерации — отрастанию утраченных частей тела после поперечной перерезки. Наблюдение за процессом регенерации осуществляется путем видеорегистрации движущейся планарии в поле зрения микроскопа. Для измерения степени регенерации [1] возникает необходимость в нормализации изображений — виртуальном приведении тела червя в «распрямленное» положение. Предлагаемый метод, основанный на построении модели тела планарии в виде жирной линии [2], включает в себя следующие шаги:

- сегментация — выделение тела планарии на изображении;
- получение медиального представления формы планарии в виде непрерывного скелета и радиальной функции;
- аппроксимация медиального представления планарии жирной кривой Безье;
- параметризация границы объекта — определение локальных координат границы относительно жирной кривой;
- распрямление оси и соответствующее преобразование границы объекта.

На основе нормализации изображения планарии оцениваются величина и скорость регенерации. Корректность и эффективность метода подтверждаются результатами вычислительных экспериментов на большой базе изображений.

Работа поддержана грантами РФФИ № 14-01-00716 и № 15-29-01135 и Госзадания Министерства образования и науки РФ, № 2912.

[1] *Тирас Х. П., Петрова О. Н., Мяжишева С. Н., Деев А. А., Асланиди К. Б.* Минимизация погрешностей морфометрии регенерирующих планарий // *Фундаментальные исследования*, 2015. № 2. Ч. 7. С. 1412–1416.

[2] *Mestetskii L. M.* Fat curves and representation of planar figures // *Comput. Graphics*, 2000. Vol. 24. No. 1. P. 9–21.

Normalization of planarian images by fat curves in biological studies

*Mestetskiy Leonid*¹★
*Tiras Kharlampiy*²

mestlm@mail.ru
tiras1950@yandex.ru

¹Moscow, Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics

²Pushchino, Pushchino State Institute of Natural Sciences

Ciliated planarians flatworms possess outstanding ability to regenerate — regrowth of lost body parts after cross-sectioning. Monitoring the process of regeneration [1] is accomplished by moving video recording planarians under the microscope. The normalization of images is necessary for measuring the level of regeneration. The normalization is a virtual “straightening” of the worm’s body.

The proposed method is based on constructing a model of a planaria body as a fat curve [2]. The fat curve is a family of circles centered on a segment of a continuous smooth curve and radii of circles are described by a smooth function of the curve point.

Solution of the problem involves the following steps:

- segmentation — extraction of the planarian body in the image;
- obtaining the medial representation of planarian shape as a continuous skeleton and radial function;
- approximation of the medial representation of planarian shape by fat Bezier curve;
- parameterization of the object’s border — defining the local coordinates of the boundaries with respect to the fat curve; and
- straightening the main axis of the skeleton and the corresponding transformation of the object’s border.

The method of evaluation of the size and speed of planarian regeneration is proposed for a normalized planarian image. Correctness and efficiency of the proposed method are confirmed by the results of computational experiments with a large database of images.

This research is funded by the Russian Foundation for Basic Research, grants 14-01-00716 and 15-29-01135 and by the Ministry of Education and Science of the Russian Federation, grant 2912.

[1] Tiras, H. P., O. N. Petrova, S. N. Myakisheva, A. A. Deev, and K. B. Aslanidi. 2015. Minimizing of morphometric errors in planarian regeneration. *Fundamental Research* 2(7):1412–1416.

[2] Mestetskii, L. 2000. Fat curves and representation of planar figures. *Comput. Graphics* 24(1):9–21.

Определение видимой области радужки классификатором локальных текстурных признаков

Соломатин Иван Андреевич^{1*}

ivansolomat@yandex.ru

*Матвеев Иван Алексеевич*²

matveev@ccas.ru

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Вычислительный центр им. А. А. Дородницына РАН

Распознавание человека по изображению радужной оболочки — актуальная задача в биометрических системах. Помимо выделения радужки как кольцевой области для повышения точности распознавания определяют области затенения (блики, веки, ресницы и др.). Задача выделения затенений радужки может быть поставлена как классификация пикселей кольцевой области на два класса: «радужка» и «затенение». В кольцевой области определяется сегмент S с минимальной дисперсией яркости, который, как правило, не содержит затенений (в данной работе этот сегмент рассматривается как часть входных данных). Обучающая выборка класса «радужка» задается по пикселям этого сегмента. Использование сектора S для обучения позволяет избежать использования обучающей выборки, размеченной вручную, т. е. построить полностью автоматический метод [1].

В работе используется классификатор на основе многомерного гауссиана, а в качестве признаков использованы яркость в точке, средняя яркость в окрестности, среднеквадратичное отклонение яркости в окрестности и пять компонент дискретного косинусного преобразования. Параметры метода оптимизированы с помощью генетического алгоритма. Результат работы классификатора подвергается морфологической постобработке, что позволяет избавиться от шума и сложностей с обработкой участков века, внешне схожих с радужкой. В качестве функции ошибки использовалась сумма относительных ошибок первого и второго рода. Метод реализован в системе Matlab. Среднее значение функции ошибки на изображениях из базы CASIA составило 0,286.

Работа поддержана грантом РФФИ № 15-01-05552.

- [1] *Соломатин И. А., Матвеев И. А.* Определение видимой области радужки классификатором локальных текстурных признаков // Машинное обучение и анализ данных, 2015 (в печати). Т. 1. №. 14. jmla.org/papers/doc/2015/JMLDA2015no14.pdf.

Detecting visible areas of iris by qualifier of local textural features

Solomatin Ivan^{1*}

ivansolomat@yandex.ru

*Matveev Ivan*²

matveev@ccas.ru

¹Moscow, Moscow Institute of Physics and Technology (State University)

²Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

A person recognition by the image of the iris is an actual problem. To increase the accuracy of recognition, usually areas of occlusion are detected in addition to locating of the iris as an annular region. The problem of occlusion detection can be set as the classification of pixels from annular region into two classes: “iris” and “occlusion.” In the annular region, the segment S with minimum dispersion of brightness is selected, which usually contains no occlusion (in this article, this segment is not calculated and supposed to be a part of the input data). The training set of class “iris” is set by local textural features of the pixels from this segment. Using sector S as training set lets to avoid using hand-marked training set and it means that the method is automatic [1].

A classifier based on a multidimensional Gaussian is used and the features are brightness, average brightness in the neighborhood, the standard deviation of the brightness in the neighborhood, and five components of the discrete cosine transform. Parameters of the algorithm are optimized by a genetic algorithm. The morphological postprocessing of the output of the classifier eliminates the noise and helps to classify areas of the eyelid which are similar to the iris. Sum of the relative errors of the first and second type is used as an error function. The method was implemented in Matlab. The average value of the error function at the images from the CASIA database was 0.286.

This research is funded by the Russian Foundation for Basic Research, grant 15-01-05552.

- [1] Solomatin, I., and Matveev I. 2015 (in press). Detecting visible areas of iris by qualifier of local textural features // *J. Mach. Learn. Data Anal.* 1(14). jmla.org/papers/doc/2015/JMLDA2015no14.pdf.

Цифровое моделирование рельефа с использованием ортогональных полиномов

*Флоринский Игорь Васильевич**

iflor@mail.ru

Панкратов Антон Николаевич

pan@impb.ru

Пуццино, Институт математических проблем биологии РАН

Математические аспекты цифрового моделирования рельефа включают интерполяцию цифровых моделей высот (ЦМВ), генерализацию и подавление шума в ЦМВ, а также расчет морфометрических характеристик на основе вычисления частных производных высоты. Эти процедуры традиционно основаны на численной обработке ЦМВ – дискретных функций двух переменных.

Авторами разработан спектрально-аналитический метод и алгоритм на основе ортогональных разложений высокого порядка с использованием многочленов Чебышёва I рода с последующим суммированием Фейера. Метод и алгоритм предназначены для аналитической обработки ЦМВ, включая глобальную аппроксимацию ЦМВ, генерализацию и подавление шума в ЦМВ, а также расчет морфометрических характеристик на основе аналитического вычисления частных производных [1].

Для тестирования метода и алгоритма использовалась ЦМВ Северных Анд, включающая 230 880 точек (матрица высот 480×481). Цифровые модели высот были восстановлены с использованием 480, 240, 120, 60 и 30 коэффициентов разложения. Первые и вторые частные производные высоты были аналитически рассчитаны по восстановленным ЦМВ. Производные были использованы для расчета моделей горизонтальной кривизны (k_h). Серия карт высоты и k_h , полученных для различного числа коэффициентов разложения, хорошо иллюстрирует генерализацию данных, подавление шума и удаление артефактов, присутствующих в исходной ЦМВ. Результаты тестирования показали высокую работоспособность предлагаемого метода и алгоритма. Они могут применяться как универсальное средство для аналитической обработки данных в цифровом моделировании рельефа.

Работа поддержана грантом РФФИ № 15-07-02484.

- [1] *Florinsky I. V., Pankratov A. N.* Digital terrain modeling with orthogonal polynomials // *J. Mach. Learn. Data Anal.*, 2015. Vol. 1. No. 12. P. 1647–1659. jmla.org/papers/doc/2015/no12/Florinsky2015Terrain.pdf.

Digital terrain modeling with orthogonal polynomials

*Florinsky Igor**

iflor@mail.ru

Pankratov Anton

pan@impb.ru

Pushchino, Institute of Mathematical Problems of Biology of the Russian Academy of Sciences

Mathematical problems of digital terrain analysis includes interpolation of digital elevation models (DEMs), DEM generalization and denoising, and computation of morphometric variables by calculation of partial derivatives of elevation. Traditionally, these procedures are based on numerical treatments of two-variable discrete functions of elevation.

The authors developed a spectral analytical method and algorithm based on high-order orthogonal expansions using the Chebyshev polynomials of the first kind with the subsequent Fejer summation. The method and algorithm are intended for DEM analytical treatment, such as DEM global approximation, generalization and denoising as well as computation of morphometric variables by analytical calculation of partial derivatives [1].

To test the method and algorithm, a DEM of the Northern Andes including 230,880 points (the elevation matrix 480×481) has been used. Digital elevation models were reconstructed with 480, 240, 120, 60, and 30 expansion coefficients. The first and second partial derivatives of elevation were analytically calculated from the reconstructed DEMs. The models of horizontal curvature (k_h) were then computed with the derivatives. A set of elevation and k_h maps related to different number of expansion coefficients illustrates well the data generalization effects, denoising, and removal of artifacts contained in the original DEM.

The test results demonstrated a good performance of the developed method and algorithm. They can be utilized as a universal tool for analytical treatment in digital terrain modeling.

This research is funded by the Russian Foundation for Basic Research, grant 15-07-02484.

- [1] Florinsky, I.V., and Pankratov A.N. 2015. Digital terrain modeling with orthogonal polynomials. *J. Mach. Learn. Data Anal.* 1(12):1647–1659. jmla.org/papers/doc/2015/no12/Florinsky2015Terrain.pdf.

Использование текстурного изображения внешней стороны пальцев в клиент-серверной системе биометрической идентификации

Чернышов Виктор Геннадьевич webcreator18@gmail.com

Местецкий Леонид Моисеевич mestlm@mail.ru

Москва, Московский государственный университет
им. М. В. Ломоносова, Факультет ВМК

Повсеместное распространение планшетов и смартфонов естественным образом стимулирует разработку новых систем биометрической идентификации на их основе. Наряду с отпечатками пальцев и радужной оболочкой глаза применение ладони для распознавания личности человека — перспективное направление в современных прикладных задачах. Характеристики мобильных устройств (вычислительная мощность, качество камеры, высокоскоростные протоколы связи) позволяют реализовать различные варианты архитектуры приложения для идентификации по ладони. Выбранная гибридная схема (расчеты как на устройстве, так и на сервере) дает свободу в размещении программных модулей и обеспечивает высокий уровень интерактивности по отношению к пользователю. Данная работа посвящена извлечению текстурных особенностей внешней стороны ладони в области костяшек пальцев и их использованию для повышения качества идентификации (ранее в публикациях распознавание опиралось на признаковое описание, извлекаемое исключительно по форме ладони). Формирование признаков основано на двумерном фильтре Габор, волновая структура которого хорошо согласуется с кожными складками костяшек. Отличительная черта предлагаемого метода — работоспособность в реальных условиях эксплуатации: анализируется видеопоток с камеры мобильного устройства, возможна съемка «с руки» (без использования штатива), поступающие кадры обрабатываются за доли секунды. Вычислительный эксперимент на массивной базе снимков подтверждает превосходство предлагаемого бимодального метода над использовавшимся ранее унимодальным [1].

Работа поддержана грантом РФФИ № 14-01-00716

- [1] *Chernyshov V.* Efficient hand detection in client-server recognition system // International Conference on Computer Vision Theory and Applications (VISAPP 2015) Proceedings. Portugal: SCITEPRESS, 2015. Vol. 2. P. 461–468.

Using finger knuckle prints in client-server biometric system

*Chernyshov Victor**

webcreator18@gmail.com

Mestetskiy Leonid

mestlm@mail.ru

Moscow, Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics

The growing market of tablets and smartphones naturally leads to developing new biometric systems based on mobile devices. Together with fingerprints and iris, utilizing hands in modern applications of identity recognition is a promising and challenging task. The characteristics of mobile devices (computational power, camera quality, and high-speed communication protocols) make it possible to implement different architecture schemes for a hand-based recognition application. Selected hybrid theme (calculations are done both on device and server) provides a flexibility in arrangement of routines and also gives a high level of interactivity for an end-up user. This work covers the problem of features extraction from finger knuckle print regions and using this description to improve identification quality (in the authors' earlier studies, only shape-based features have been employed). The backbone of the proposed method is a two-dimensional Gabor filter chosen because of similarity between its wave-like structure and skin folds in knuckle regions. The key feature of the method is an efficiency in real-life conditions: video sequence from a mobile camera is analyzed, capturing may be done "from hand" (without using tripod), and captured frames are processed for a fraction of a second. The superiority of introduced bimodal method over the previous unimodal is proved by numerical experiment on a massive hand image dataset [1].

This research is funded by the Russian Foundation for Basic Research, grant 14-01-00716

- [1] Chernyshov, V. 2015. Efficient hand detection in client-server recognition system. *International Conference on Computer Vision Theory and Applications (VISAPP 2015) Proceedings*. Portugal: SCITEPRESS. 2:461–468.

Оценивание характера последствия случайных точечных процессов методами многомасштабного корреляционного анализа

*Анциперов Вячеслав Евгеньевич** antciperov@cplire.ru

Москва, Институт радиотехники и электроники
им. В. А. Котельникова РАН

Обсуждается применение методов многомасштабного корреляционного анализа к проблеме оценивания характеристик широкополосных импульсных процессов с независимыми интервалами между импульсами (точечных процессов с ограниченным последствием) [1]. Показано, что для усредненных многомасштабных (конусных) представлений удастся, если использовать технику аналитических спектров и, соответственно, методы теории функций комплексного переменного, существенно упростить их вид. Это обстоятельство существенно облегчает анализ зависимости характерных особенностей представлений от параметров потока импульсов. К последним относятся, например, средняя интенсивность, ее вариабельность, характер последствия и др.

На основе выявленных аналитических зависимостей предложено построение соответствующих оценок по выборочным данным — реализациям импульсных процессов. Среди случаев, в которых удастся провести аналитические вычисления до квадратур, рассматриваются случай стационарного потока и случай резкого изменения стационарности. Эти случаи моделируют две крайние ситуации — квазистационарного поведения процесса и переходной режим.

Работа поддержана грантом РФФИ № 15-07-04378-а.

- [1] *Анциперов В. Е.* Оценивание характера последствия случайных точечных процессов методами многомасштабного корреляционного анализа // Журнал радиоэлектроники: электронный журнал, 2015. № 6. <http://jre.cplire.ru/jre/jun15/8/text.pdf>.

Estimation of renewal point process aftereffect characteristics on the bases of multiscale correlation analysis methods.

*Antsiperov Viacheslav**

antsiperov@cplire.ru

Moscow, Kotel'nikov Institute of Radioengineering and Electronics
of the Russian Academy of Sciences

The application of multiscale correlation analysis to the problem of evaluating the characteristics of wideband pulse processes with independent intervals (point processes with limited aftereffects — renewal processes) is discussed [1]. It is shown that for the averaged multiscale (cone kernel) representations, it is possible, using the technique of analytic spectra and, accordingly, the methods of the theory of functions of a complex variable, to simplify significantly their appearance. The latter greatly facilitates the dependence of representation characteristics on the process parameters such as average intensity, its variability, character of aftereffects, etc.

Basing on the analytical dependences found, the paper suggests a construction of the corresponding estimates formed by sampled data — pulse processes realizations. Among the cases in which it is possible to carry out the analytical calculation to the final quadrature, the case of a stationary flow and the case of sudden stationarity changes are considered.

This research is funded by the Russian Foundation for Basic Research, grant 15-07-04378-a.

- [1] Antsiperov, V. 2015. Estimation of renewal point process aftereffect characteristics on the bases of multiscale correlation analysis methods. *J. Radio Electronics* 6. <http://jre.cplire.ru/jre/jun15/8/text.pdf>.

Оценивание параметров радиоимпульсов методом максимального правдоподобия

Власова Ксения Валерьевна^{1*} p_ksenia@mail.ru

*Пахотин Валерий Анатольевич*² VPakhotin@kantiana.ru

*Клионский Дмитрий Михайлович*³ klio2003@list.ru

*Каплун Дмитрий Ильич*³ dikaplun@etu.ru

¹Калининград, Балтийская государственная академия
рыбопромышленного флота

²Калининград, Балтийский Федеральный университет им. И. Канта

³Санкт-Петербург, Санкт-Петербургский государственный
электротехнический университет «ЛЭТИ»

Настоящая статья направлена на решение задачи оценки параметров двух неортогональных радиоимпульсов (перекрытие спектров которых не удовлетворяет критерию Рэлея). Получение решения методом максимального правдоподобия в данных условиях затруднительно, поскольку в области неортогональности логарифм функции правдоподобия имеет сложную структуру (множество локальных минимумов) и результат его минимизации неоднозначен. В работе показано, что возможна модернизация метода максимального правдоподобия за счет преобразования функционала правдоподобия с учетом системы уравнений правдоподобия, при этом решение получается единственным. Модельные исследования показали, что возможно разрешение двух неортогональных радиоимпульсов и оценка их параметров до значения коэффициента корреляции 0,9. При этом для сохранения дисперсии оценок параметров на прежнем уровне требуется повысить отношение сигнал/шум на 7 дБ. В работе приводится обоснование возможности оценки дисперсии шума непосредственно по выборке данных. Она совпадает со значением преобразованного функционала правдоподобия в точке глобального минимума [1].

Работа поддержана грантом РФФИ № 15-37-50902 и контрактом № 02.G25.31.0058 от 12.02.2013 (Минобрнауки России).

- [1] *Власова К.В., Пахотин В.А., Клионский Д.М., Каплун Д.И.* Оценивание параметров радиоимпульсов методом максимального правдоподобия // Машинное обучение и анализ данных, 2015 (в печати). Т. 1. №. 14. jmla.org/papers/doc/2015/JMLDA2015no14.pdf.

Estimation of radio impulse parameters using the maximum likelihood method

Vlasova Ksenia^{1*}

p_ksenia@mail.ru

Pachotin Valeriy²

VPachotin@kantiana.ru

Klionskiy Dmitriy³

klio2003@list.ru

Kaplun Dmitriy³

dikaplun@etu.ru

¹Kaliningrad, Baltic Fishing Fleet State Academy

²Kaliningrad, Immanuel Kant Baltic Federal University

³Saint Petersburg, Saint Petersburg Electrotechnical University "LETI"

The present paper discusses the solution of a parameter estimation problem for two nonorthogonal radio impulses (radio impulses are received simultaneously and their nonorthogonality is established in the frequency domain). Radio impulses are considered to be nonorthogonal if their spectral overlap does not satisfy the Rayleigh criterion. Application of spectral analysis and the maximum likelihood method do not allow to obtain a solution since the logarithm of the likelihood function has a complex structure (many local minima) in the nonorthogonality area and its minimization is ambiguous. It is shown that it is possible to update the maximum likelihood method due to the transform of the likelihood functional using likelihood equations. As a result, a unique solution can be obtained and it is shown that parameter estimate variances depend on the correlation coefficient between signals and signal-to-noise ratio. Model investigations have shown that it is possible to provide time resolution of two nonorthogonal radio impulses and to estimate their parameters if the correlation coefficient does not exceed 0.9. To preserve parameter estimate variances, it is necessary to increase the signal-to-noise ratio by 7 dB. Also, noise variance estimation was obtained using the given data set [1].

The paper is supported by the Russian Foundation for Basic Research (grant 15-37-50902) and by the Russian Ministry of Education and Science (contract 02.G25.31.0058 dated 12.02.2013).

- [1] Vlasova, K. V., Pachotin V. A., Klionskiy D. M., and Kaplun D. I. 2015 (in press). Estimation of radio impulse parameters using the maximum likelihood method. *J. Mach. Learn. Data Anal.* 1(14). jmla.org/papers/doc/2015/JMLDA2015no14.pdf.

Перспективы использования математических методов распознавания образов в технологиях встраивания цифровых водяных знаков в аудиосигнал

*Жарких Александр Александрович** zharkikh090107@mail.ru
Мурманск, ФГБОУ ВПО МГТУ

Цель данного сообщения — презентация монографии [1]. В докладе будут представлены алгоритмы встраивания цифровых водяных знаков (ЦВЗ) в аудиосигнал с указанием их связи с классическими методами математики и математической теорией распознавания образов.

Авторы [1] предприняли попытку систематического изложения различных аспектов технологий встраивания ЦВЗ в аудиосигнал. В нескольких разделах авторы описывают общие принципы стеганографии, выделяют встраивание ЦВЗ как один из методов стеганографии, анализируют области применимости ЦВЗ. Представлены физические и психофизиологические аспекты выбора аудиосигнала в качестве контейнера и обоснован выбор аудиоформата wav. Авторы предлагают три группы методов встраивания ЦВЗ, каждая из которых обладает своими преимуществами и недостатками. Отражены взгляды авторов на перспективы развития.

Отметим следующие направления перспективного развития технологий встраивания ЦВЗ в аудиосигналы: развитие линейных, нелинейных и помехоустойчивых методов, исследованных в монографии [1]; развитие методов встраивания ЦВЗ в речевой сигнал путем вариации просодических параметров речи; развитие стегоанализа, основанного на методах теории вероятности и математической статистики.

Отмеченные перспективные задачи могут быть качественно решены лишь с использованием алгебраической и статистической теории распознавания образов.

- [1] *Гурин А. В., Жарких А. А., Пластунов В. Ю.* Технологии встраивания цифровых водяных знаков в аудиосигнал / Под общ. ред. А. А. Жарких. — М.: Горячая линия Телеком, 2015. 116 с. http://www.techbook.ru/book.php?id_book=778.

Perspectives of uses of mathematical methods of image identification in technologies of embedding of digital water marks in an audio signal

Zharkikh Aleksandr*

zharkikh090107@mail.ru

Murmansk, FSEI HPE MSTU

The purpose of this message is to present the monograph [1]. In the report, algorithms of embedding of the digital water marks (DWM) will be provided to an audio signal with specifying their connection with classical methods of mathematics and mathematical theory of image identification.

The authors of [1] made an attempt of systematic presentation of different aspects of technologies of embedding of DWM in an audio signal. In several sections, the authors describe the general principles of a steganography, select embedding of DWM as one of steganography methods, and analyze the areas of applicability of DWM. Physical and psychophysiological aspects of a choice of an audio signal as the container are provided and wav audio format choice is shown to be reasonable. The authors of [1] suggest three groups of DWM embedding methods, each with the advantages and shortcomings. The book reflects the authors' understanding of development perspectives.

Let mark the following perspective directions of development technologies of embedding DWM in audio signals: development of the linear, nonlinear, and noiseproof methods researched in the monograph [1]; development of methods of embedding DWM in a voice signal by a variation of prosodic parameters of the speech; and development of the stegoanalyse based on the methods of probability theory and mathematical statistics.

The outlined perspective tasks can be solved qualitatively only with use of the algebraic and statistical theory of image identification.

- [1] Gurin, A. V., Zharkikh A. A., and Plastunov V. U. 2015. *Technologies of embedding digital watermarks in audio signal*. Ed. A. A. Zharkikh. Moscow: Telecom. 116 p. http://www.techbook.ru/book.php?id_book=778.

Анализ возможностей использования технологий data mining для изучения сложных сигналов и систем

Жарких Александр Александрович zharkikh090107@mail.ru
*Слащев Алексей Александрович** chronoexplosion@gmail.com
Мурманск, ФГБОУ ВПО МГТУ

Представлены базовые элементы технологии data mining, основное назначение которой — извлечение новых знаний из некоторого объема данных. Также рассмотрены традиционные понятия сложного сигнала и сложной системы. В качестве основного параметра сложного сигнала предлагается использовать базы сигнала. Основным показателем сложной системы является наличие в ней следующих элементов: блоков ввода и вывода информации, управления информацией и блока памяти, а также совокупности блоков вычисления и преобразования сигналов. В качестве возможных областей применения технологий data mining рассматриваются задачи оценивания формы и параметров сложных сигналов, а также задачи определения вида нелинейной функции в блоках вычисления и преобразования системы.

Технологии data mining могут быть использованы для изучения и проектирования радиоэлектронных систем с расширением спектра сигналов, автоматических систем анализа и синтеза речи, компьютерных композиторов и различных объектов из других предметных областей. Для их использования необходимо провести такое количество предварительных экспериментов, которое позволит создать обучающую выборку, а также развитие математических моделей, алгоритмов, методов и программных средств для изучения и проектирования соответствующих объектов.

В частности, использование data mining для изучения свойств сигналов и систем требует создание базы данных о параметрах этих сигналов и систем. Также необходимо развитие линейных и нелинейных моделей математики и соответствующих программных средств.

Analysis of possibilities in using data mining technologies to study complex signals and systems

Zharkikh Aleksandr

zharkikh090107@mail.ru

*Slashchev Aleksey**

chronoexplosion@gmail.com

Murmansk, FSEI HPE MSTU

Basic elements of the data mining technology are provided. The basic purpose of these technologies is an extraction of new knowledge from some data volume. Traditional concepts of a complex signal and complex system are also considered. Use of basis of a signal is suggested as the key parameter of a complex signal. The main index of a complex system is determined by the following elements in it: information input and output units, information control unit, memory unit, and a set of signal computation and conversion units. Applications of the data mining technologies tasks for estimation of the form and parameters of complex signals and also the tasks for type determination of nonlinear function in the system conversion unit are considered.

The data mining technologies can be used for study and design of radioelectronic systems with extension of a range of signals, automatic systems of speech analysis and synthesis, computer composers and different objects from other data fields. To use them, at first, it is necessary to carry out such amount of preliminary experiments which will allow to create the learning selection. The second preliminary requirement consists in development of mathematical models, algorithms, methods, and software for study and design of the appropriate objects.

In particular, the use of data mining for signal properties and systems studies requires creation of a database containing parameters of these signals and systems. Also, development of the linear and nonlinear mathematics models and the appropriate software are necessary.

Разработка и исследование свойств системы прогнозирования объемов спроса на грузовые железнодорожные перевозки

*Рудаков Константин Владимирович*¹ rudakov@ccas.ru

*Каширин Даниил Олегович*¹ kashirin@forecsys.ru

Стрижов Вадим Викторович^{1*} strijov@ccas.ru

*Корчагин Александр Дмитриевич*²

¹Москва, Вычислительный центр им. А. А. Дородницына РАН

²Москва, Центр инновационного развития ОАО «РЖД»

Описаны результаты исследований по созданию системы прогнозирования, направленной на повышение качества управления грузовыми железнодорожными перевозками. Задача прогнозирования объемов спроса на грузовые перевозки была поставлена для оперативного планирования перевозок по историческим данным. Данные содержат информацию о дате погрузки, станции отправления, станции назначения, количестве вагонов, которые прошли по маршруту, код груза, род вагонов, суммарный вес груза в тоннах и признак маршрутной отправки. Модель прогнозирования должна формировать прогноз спроса на отправление/погрузку грузов с разложением по группам грузов, родам вагонов и станциям отправления/назначения.

Для решения данной задачи был поставлен и решен ряд подзадач, связанных со спецификой прикладной области:

- предложен алгоритм непараметрического прогнозирования; прогнозы выполняются по парам станций, с учетом топологии сети [1];
- поставлена задача выявления и исследования экзогенных факторов и характера их влияния на объемы грузовых железнодорожных перевозок;
- предложен способ согласования независимых прогнозов объемов грузоперевозок.

Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации (соглашение № RFMEFI60414X0041)

- [1] Газизуллина Р. К., Стенина М. М., Стрижов В. В. Прогнозирование объемов железнодорожных грузоперевозок по парам веток // Системы и средства информатики, 2015. Т. 25. Вып. 1. С. 142–154.

Design and analysis of a forecasting system for rail freight transportation demand

*Rudakov Konstantin*¹

rudakov@ccas.ru

*Kashirin Daniil*¹

kashirin@forecsys.ru

Strijov Vadim^{1*}

strijov@ccas.ru

*Korchagin Alexander*²

¹Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

²Moscow, RZD Innovative Development Centre

The results of research aimed to create a forecasting system are reported. The goal is to increase the quality of railway freight transportation management. The forecasting problem has been formulated for operative (short-term) transportation planning. The system outputs daily, weekly, or monthly forecasts in various ranges up to one year. Historical data contain information on shipping date, origin and destination stations, number of shipped wagons, freight type, wagon type, total weight of shipped freight, and a transfer number. The system must output forecasts with specification of freight type, wagon type, and origin/destination stations.

To achieve this goal, the authors considered a number of subproblems, associated with particularities of the application:

- a nonparametric forecasting algorithm has been proposed. The forecasts are obtained for pairs of stations with respect to the rail network topology [1];
- to further increase forecasting quality, possible exogenous factors, which influence transportation demand, have been investigated. This includes defining most influential factors and the way they influence forecasted time series; and
- a method of reconciling independent forecasts with different specifications of freight type, wagon type, and origin/destination stations has been proposed.

The research was supported by the Ministry of Education and Science of the Russian Federation, RFMEFI60414X0041.

- [1] Gazizullina, R. K., Medvednikova M. M., and Strijov V. V. Capacity of railway cargo transportation forecasting. 2015. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(1):142–154.

Восстановление пропущенных данных на ультразвуковых дефектограммах рельсового пути на основе локально-глобального выравнивания

Чепрасов Дмитрий Николаевич^{1*} tcheprasov.mitya@yandex.ru

*Маленичев Антон Александрович*¹ malenichev@mail.ru

*Сулимова Валентина Вячеславовна*¹ vsulimova@yandex.ru

*Красоткина Ольга Вячеславовна*¹ o.v.krasotkina@yandex.ru

*Моттль Вадим Вячеславович*² vmottl@yandex.ru

*Марков Анатолий Аркадиевич*³ amarkovspb@gmail.com

¹Тула, Тульский государственный университет

²Москва, Московский физико-технический институт (ГУ)

³Санкт-Петербург, ОАО «Радиоавионика»

Работа посвящена актуальной проблеме автоматического восстановления пропущенных данных при ультразвуковом контроле рельсового пути, препятствующих своевременному обнаружению дефектов рельсов. Задача восстановления пропущенных данных формулируется в данной работе как задача поиска соответствующей области на ультразвуковой дефектограмме предыдущего прохода по тому же участку. Авторы усовершенствуют предложенный ими ранее метод, делая его более точным и существенно более быстрым. Это достигается за счет двух основных нововведений: во-первых, применения специальной трехконной модели для быстрой примерной локализации опорных зон болтового соединения и, во-вторых, применения специальной локально-глобальной процедуры парного выравнивания фрагментов дефектограмм для уточнения их положения. Предлагаемый подход позволяет определить на дефектограмме предыдущего прохода положение области, соответствующей пропуску на текущем проходе, с точностью около 3,5 см, что обеспечивает возможность успешного восстановления по ней пропущенных данных [1].

Работа поддержана грантом РФФИ № 15-37-50751.

- [1] *Чепрасов Д. Н., Маленичев А. А., Сулимова В. В., Красоткина О. В., Моттль В. В., Марков А. А.* Восстановление пропущенных данных на ультразвуковых дефектограммах рельсового пути на основе локально-глобального выравнивания // Машинное обучение и анализ данных, 2015 (в печати). Т. 1. №.12. jmla.org/papers/doc/2015/JMLDA2015no12.pdf

Recovering missing data on ultrasonic rail defectograms via semiglobal warping

*Cheprasov Dmitry*¹★

tcheprasov.mitya@yandex.ru

*Malenichev Anton*¹

malenichev@mail.ru

*Sulimova Valentina*¹

vsulimova@yandex.ru

*Krasotkina Olga*¹

o.v.krasotkina@yandex.ru

*Mottl Vadim*²

vmottl@yandex.ru

*Markov Anatoly*³

amarkovspb@gmail.com

¹Tula, Tula State University

²Moscow, Moscow Institute of Physics and Technology (State University)

³Saint Petersburg, Radioavionica Corp.

The damages on rails are exceptionally dangerous for the operation of rail traffic. In this connection, during the last years, mobile ultrasonic defectoscopes are actively introduced to inspect rail tracks for damages. These devices register reflected ultrasonic signals at defectograms.

One of the actual problems of ultrasonic railway inspection is presence of missing data, i. e., uncontrolled regions, taking place due to a bad acoustic contact.

The paper deals with the problem of automatic recovering such missing data on defectograms. Here, the problem of missing data recovery is formulated as the problem of finding the respective region at the defectogram from the previous ultrasonic inspection.

In this work, the present authors update their previous method, making it more accurate and appreciably more fast one. It is possible due to applying two main innovations: (i) introducing 3-windows model for fast approximate detection of positions of support bolt-on joint areas and (ii) applying special semiglobal warping procedure for defectograms. The proposed approach allows to detect the area of interest at the previous defectogram accurate within 3.5 cm and, so, it provides the possibility to successful automatic data recovery from the previous ultrasonic defectogram [1].

This research is funded by the Russian Foundation for Basic Research, grant 15-37-50751.

- [1] Cheprasov, D., Malenichev A., Sulimova V., Krasotkina O., Mottl V., Markov A. 2015 (in press). Recovering missing data on ultrasonic rail defectograms via semiglobal warping. *J. Mach. Learn. Data Anal.* 1(12). jmla.org/papers/doc/2015/JMLDA2015no12.pdf.

Разработка подсистемы маршрутизации пациентов в рамках интеллектуальной пациент-ориентированной системы управления крупной клиникой

<i>Дорофеев Юлия Александровна</i> ^{1*}	dorofeyuk_julia@mail.ru
<i>Глазман Леонид Юрьевич</i> ²	lglasman@nsi.ru
<i>Дорофеев Александр Александрович</i> ^{1,3}	daa2@mail.ru
<i>Мандель Александр Соломонович</i> ¹	almandel@yandex.ru
<i>Потапов Александр Александрович</i> ²	APotapov@nsi.ru
<i>Чернявский Александр Леонидович</i> ¹	achern@ipu.ru
<i>Шиманский Вадим Николаевич</i> ²	ailyash@nsi.ru
<i>Шифрин Михаил Абрамович</i> ²	shifrin@nsi.ru

¹Москва, Институт проблем управления РАН

²Москва, НИЦ ИНХ МЗ

³Москва, Национальный исследовательский университет «Высшая школа экономики»

Статья посвящена разработке интеллектуальной информационной подсистемы маршрутизации, предназначенной для управления потоками пациентов при оказании лечебно-диагностических услуг в медицинском учреждении. Эта подсистема — один из важных блоков пациент-ориентированной интеллектуальной системы управления лечебно-диагностическими процессами в медицинском учреждении (на примере клиники нейрохирургического профиля). По результатам проведенной коллективной многовариантной экспертизы, результатам интеллектуальной обработки экспертных оценок и используя имеющийся зарубежный опыт в этом направлении, были сформулированы основные положения концепции адаптивной интеллектуальной информационной подсистемы маршрутизации [1].

Работа выполнена при частичной финансовой поддержке РФФИ, гранты №№ 13-07-12201, 13-07-00992, 14-07-00463 и 15-07-06713.

- [1] *Дорофеев А. А., Потапов А. А., Чернявский А. Л., Шифрин М. А.* Создание пациент-ориентированной интеллектуальной системы управления лечебно-диагностическими процессами в крупном медицинском учреждении // Управление развитием крупномасштабных систем (MLSD'2013): Мат-лы 7-й междунар. конф. М: ИПУ РАН, 2013. Том 2. С. 402–405.

Development of subsystem routing patients through intelligent patient-oriented system for the management of large clinic

<i>Dorofeyuk Yuliya</i> ¹ ★	dorofeyuk_julia@mail.ru
<i>Glazman Leonid</i> ²	lglasman@nsi.ru
<i>Dorofeyuk Alexander</i> ^{1,3}	daa2@mail.ru
<i>Mandel Alexander</i> ¹	almandel@yandex.ru
<i>Potapov Alexander</i> ²	APotapov@nsi.ru
<i>Chernyavskiy Alexander</i> ¹	achern@ipu.ru
<i>Shimansky Vadim</i> ²	ailyash@nsi.ru
<i>Shifrin Mihail</i> ²	shifrin@nsi.ru

¹Moscow, Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences

²Moscow, SIC NSI MHC RF

³Moscow, National Research University “Higher School of Economics”

Paper is devoted to the development of intelligent information routing subsystem designed to manage the flow of patients in the provision of medical diagnostic services in a medical facility. This subsystem is one of the important blocks of a patient-oriented intelligent control of diagnostic and treatment processes systems in a medical facility (on the example of the neurosurgical clinic). According to the results of collective multivariate examination, the results of data mining expert assessments, and using the foreign experience in this direction, the basic provisions of the adaptive intelligent information routing subsystem concept have been formulated [1].

The research is performed under partial financial support of the Russian Foundation for Basic Research, grants 13-07-12201, 13-07-00992, 14-07-00463, and 15-07-06713.

- [1] Dorofeyuk, A., Potapov A., Chernyavsky A., and Shifrin M. 2013. The creation of patient-oriented intelligent control systems diagnostic and treatment processes in a medical institution. *Managing the Development of Large-Scale Systems (MLSD'2013): 7th International Conference Proceedings*. Moscow: ICS RAS. 2:17–22.

Локализация источников биомагнитной активности мозга

*Махортых Сергей Александрович**

makh@impb.ru

Лызко Екатерина Викторовна

lyzko@yandex.ru

Пушино, Институт математических проблем биологии РАН

Предлагается метод классификации типов активности головного мозга для данных магнитоэнцефалографии (МЭГ). Источники аномальной активности коры мозга локализуются на основе применения обобщенного спектрального анализа в пространстве коэффициентов Фурье разложений полученных сигналов по настраиваемым ортогональным базисам. Обсуждаются основные принципы метода и результаты его применения к реальным записям МЭГ. Результаты используются также для функционального картирования мозга в нормальных и патологических состояниях.

Интервалы аномальной и нормальной активности могут быть легко идентифицированы визуально, если они отличаются по средней амплитуде сигнала. В противном случае визуальный анализ не является надежным для идентификации интервалов, соответствующих различным режимам деятельности. В этом случае тип активности определяется с помощью анализа пространственной структуры поля, которая сильно коррелирует с количеством, местоположением и ориентацией токовых источников, ответственных за генерацию магнитного поля.

Предложенный подход был протестирован на данных, полученных на контрольной группе здоровых испытуемых и группе пациентов, страдающих болезнью Паркинсона.

Предварительная обработка данных с помощью обобщенного спектрального метода (Дедус, Махортых и др.) позволяет удалять шум, выделять необходимый сигнал, и во многих случаях определяет успех решения задачи классификации и локализации источников активности мозга [1].

Работа поддержана грантами РФФИ 14-07-00419 и 13-01-00340.

- [1] *Makhortykh S. A.* Sources localization for brain biomagnetic activity // *Rev. Appl. Phys.*, 2014. Vol. 3. P. 25–28. <http://www.seipub.org/rap/paperInfo.aspx?ID=17562>.

Sources localization for brain biomagnetic activity

*Makhortykh Sergey**

makh@impb.ru

Lyzhko Ekaterina

lyzko@yandex.ru

Pushchino, Institute of Mathematical Problems of Biology of the Russian Academy of Sciences

A method for classifying the types of brain activity in magnetoencephalographic (MEG) signals is proposed. The sources of abnormal cortical activity are localized by performing a generalized spectral analysis in the space of Fourier coefficients of the expansions of recorded signals in adaptive orthogonal bases. The basic principles of the method are discussed and the results of its application to actual MEG records are presented for functional brain mapping in normal and pathological states.

Intervals of abnormal and normal activity can be easily identified visually when they differ by the mean amplitude of the signal. Otherwise, the dominant type of activity can be determined by analyzing the spatial field pattern which correlates with the number, location, and orientation of the current sources responsible for magnetic field generation.

The proposed approach was validated against data derived from a reference group of healthy individuals and a group of patients suffering from Parkinson's syndrome.

Data preprocessing by means of generalized spectral method (Dedus, Makhortykh, *et al.*) allows to remove noise, highlight the desired signal, and, in many cases, determines the success of the solution. It uses a set of orthogonal bases so that spectral description can be scaled down to fit a particular class of signals. In many cases, correct source localization and model parameters estimates can be obtained only after projection of the signal on the proper basic function. Also, expansion coefficients can be used as features for the system control and signal type recognition [1].

This research is funded by the Russian Foundation for Basic Research, grants 14-07-00419 and 13-01-00340.

[1] Makhortykh, S. 2014. Sources localization for brain biomagnetic activity. *Rev. Appl. Phys.* 3:25–28. <http://www.seipub.org/rap/paperInfo.aspx?ID=17562>.

Анализ стадий анестезии на основе оценки энтропии электроэнцефалограммы

*Немирко Анатолий Павлович** apn-bs@yandex.ru
Калиниченко Александр Николаевич ank-bs@yandex.ru
Манило Людмила Алексеевна lmanilo@yandex.ru

Санкт-Петербург, Санкт-Петербургский государственный
электротехнический университет «ЛЭТИ»

Для контроля глубины анестезии при хирургической операции по электроэнцефалограмме (ЭЭГ) разработан алгоритм, основанный на совместном использовании анализа хаотичности сигнала, анализа частотных характеристик сигнала и анализа специфических изменений ЭЭГ, возникающих в состоянии анестезии.

Исследование осуществлялось с использованием набора реальных записей ЭЭГ, снятых в ходе проведения хирургических операций. Для анализа хаотичности сигнала предложено использовать аппроксимированную энтропию, позволяющую получить достоверные оценки по относительно коротким выборкам.

Определены оптимальные параметры процедуры вычисления аппроксимированной энтропии с точки зрения дифференциации различных состояний анестезии. Установлено, что наиболее информативным частотным параметром является относительное содержание низкочастотных составляющих сигнала в диапазоне дельта-ритма ЭЭГ. Предложен количественный показатель эффекта «всплеск-подавление», характерного для состояния глубокой анестезии.

Разработаны нелинейные шкалы и решающие правила, позволяющие получить интегральную оценку глубины анестезии, качество которой не уступает лучшим мировым результатам, что позволяет рассматривать предложенный метод как основу для создания практического алгоритма анализа ЭЭГ в мониторах глубины анестезии [1].

Работа поддержана грантами РФФИ №№ 15-07-01790, 13-01-00540 и 13-07-00721.

- [1] *Kalinichenko A. N., Manilo L. A., Nemirko A. P.* Analysis of anesthesia stages based on the EEG entropy estimation // *Pattern Recogn. Image Anal.*, 2015. Vol. 25. No. 4.

Analysis of anesthesia stages based on the electroencephalography entropy estimation

*Nemirko Anatolii**

apn-bs@yandex.ru

Kalinichenko Alexandr

ank-bs@yandex.ru

Manilo Ludmila

lmanilo@yandex.ru

Saint Petersburg, Saint Petersburg Electrotechnical University "LETI"

For the control of anesthesia depth in the course of surgical operations using electroencephalography (EEG), an algorithm based upon joint use of signal randomness analysis of approximate entropy, analysis of the signal spectrum, and analysis of EEG signal changes that are specific to the condition of deep anesthesia has been developed.

A data set containing one-channel EEG records obtained in the course of surgical operations was used for the examination. Overall duration of the EEG records was above 140 h. The approximate entropy was suggested as an estimate of the EEG signal randomness taking into account the possibility of consistent estimates obtained from the comparatively short data fragments. Optimal values of the approximate entropy calculation procedure parameters were determined from the point of view of various anesthesia stages differentiation. It was established that relative part of the low-frequency components belonging to the delta rhythm range represents the most indicative frequency parameter. Quantitative estimation of the so-called "burst-suppression" phenomenon that is specific for the state of deep anesthesia was suggested.

Nonlinear scales and decision-making rules were developed to provide an integral index characterizing patients' anesthesia depth. Comparison of the obtained results with the indications of commercially manufactured anesthesia depth monitor BIS A-2000XP (Aspect Medical) shows high degree of correlation that means that the suggested method can be used as a basement for the development of new EEG analysis algorithm for the anesthesia depth assessment devices [1].

This research is funded by the Russian Foundation for Basic Research, grants 15-07-01790, 13-01-00540, and 13-07-00721.

- [1] Kalinichenko, A. N., Manilo L. A., and Nemirko A. P. 2015. Analysis of anesthesia stages based on the EEG entropy estimation. *Pattern Recogn. Image Anal.* 25(4).

Конформационный анализ структурных мотивов типа α - α -уголок в вычислительном эксперименте молекулярной динамики

Панкратов Антон Николаевич^{1*} pan@impb.ru
*Руднев Владимир Ремович*¹ volodyarv@mail.ru
*Куликова Людмила Ивановна*¹ likulikova@mail.ru
*Дедус Флоренц Федорович*¹ fdedus@mail.ru
*Тихонов Дмитрий Анатольевич*¹ dmitry.tikhonov@gmail.com
*Ефимов Александр Васильевич*² efimov@protres.ru

¹Пушино, Институт математических проблем биологии РАН

²Пушино, Институт белка РАН

В настоящей работе решается задача распознавания структурных мотивов белков на примере α - α -уголков на основе аналитического описания основной цепи белковой глобулы и спектрального метода распознавания повторов.

Для набора структур были получены траектории молекулярной динамики для изучения структурной устойчивости. Проведен анализ найденных структур по конформационному шаблону, охарактеризованному предельными значениями углов на карте Рамачандрана. Конформационный анализ подтвердил устойчивость структур; позволил разделить класс структур с короткой перетяжкой и более общий случай; показал, что части длинных α -спиралей, выходящих за пределы гидрофобного ядра, менее устойчивы по сравнению с частями, входящими в гидрофобное ядро уголка, что проявлялось в частичном расплетании и самосборке спиралей при сохранении общей конформации уголка; позволил определить области возможных флуктуаций конформации структуры и статистически обосновать выбор границ областей параметров метода распознавания [1].

Работа поддержана РФФИ, проекты №№ 13-01-00340, 13-04-00150, 14-07-31196, 14-07-00924, 15-29-07063.

- [1] Руднев В. Р., Панкратов А. Н., Куликова Л. И., Дедус Ф. Ф., Тихонов Д. А., Ефимов А. В. Конформационный анализ структурных мотивов типа α - α -уголок в вычислительном эксперименте молекулярной динамики // Математическая биология и биоинформатика, 2014. Т. 9. № 2. С. 575–584. www.matbio.org/2014/Rudnev_9_575.pdf

Conformational analysis of structural motifs of α - α -corner in the computational experiment of molecular dynamics

Pankratov Anton^{1*}

pan@impb.ru

*Rudnev Vladimir*¹

volodyarv@mail.ru

*Kulikova Liudmila*¹

likulikova@mail.ru

*Dedus Florents*¹

fdedus@mail.ru

*Tikhonov Dmitry*¹

dmitry.tikhonov@gmail.com

*Efimov Alexander*²

efimov@protres.ru

¹Pushchino, Institute of Mathematical Problems of Biology of the Russian Academy of Sciences

²Pushchino, Institute of Protein Research of the Russian Academy of Sciences

In this paper, the authors solve the problem of recognition of structural motifs of proteins on the example of α - α -corners based on the analytical description of the main chain of the protein globule and spectral recognition method.

For a set of structures, there were obtained trajectories of molecular dynamics to study structural stability. The analysis of the found structures is performed by the conformational pattern, which is characterized by the limited values on the angles on the Ramachandran plot. Conformational analysis confirmed the stability of the structure; allowed to split a class of structures with a short constriction and the more general case; showed that long helices beyond the hydrophobic core are less stable compared with the ones included in the hydrophobic core of corner that manifested itself in the partial unwinding of the spirals and self-assembly while maintaining the common conformation of corner; and allowed to determine the range of possible fluctuations in the conformation of the structure and statistically justify the choice of the boundaries of parameters of recognition method [1].

This research is funded by Russian Foundation for Basic Research, grants 13-01-00340, 13-04-00150, 14-07-31196, 14-07-00924, 15-29-07063.

- [1] Rudnev, V., Pankratov A., Kulikova L., Dedus F., Tikhonov D., and Efimov A. 2014. Conformational analysis of structural motifs of α - α -corner in the computational experiment of molecular dynamics. *Math. Biol. Bioinf.* 9(2):575–584. www.matbio.org/2014/Rudnev_9_575.pdf

Методы классификации и распознавания образов для анализа квазипериодических биосигналов в задачах медицинской диагностики (на примере пульсового сигнала лучевой артерии)

Покровская Ирина Вячеславовна^{1*} ivp750@mail.ru
*Гучук Владимир Всеволодович*¹ polma@bk.ru
*Десова Аэлита Арсеньевна*¹ adesova@mail.ru
Дорофеев Александр Александрович^{1,2} daa2@mail.ru
*Дорофеев Юлия Александровна*¹ dorofeyuk_julia@mail.ru

¹Москва, Институт проблем управления РАН

²Москва, Национальный исследовательский университет «Высшая школа экономики»

Описывается набор методов классификации и распознавания образов, предназначенный для анализа квазипериодических биосигналов при решении задач медицинской диагностики, который был реализован как аппаратно-программный комплекс, обеспечивающий съем, кодирование, запись в базу данных, визуализацию и интеллектуальный анализ многопараметрических характеристик квазипериодического биосигнала в режиме реального времени. Рассмотрение ведется на примере обработки пульсового сигнала лучевой артерии. Аппаратно-программный комплекс включает в себя: датчик сигнала, устройство оцифровки аналогового сигнала и программное обеспечение, разработанное авторами на базе CodeGear RAD Studio. Аппаратно-программный комплекс показал свою высокую эффективность при решении прикладных задач медицинской диагностики, в том числе на базе результатов анализа ритмической структуры и формы пульсового сигнала лучевой артерии была решена задача ранней диагностики артериальной гипертензии у детей и подростков [1].

Работа выполнена при частичной финансовой поддержке РФФИ, гранты №№ 13-07-00229, 13-07-00992, 13-07-12201 и 14-07-00463.

- [1] *Гучук В. В., Десова А. А., Дорофеев А. А., Киселева Н. Е.* Аппаратно-программное обеспечение методов анализа квазипериодических биосигналов (на примере пульсового сигнала лучевой артерии) // Датчики и системы, 2014. № 8. С. 17–22.

Methods of classification and pattern recognition for the analysis of quasi-periodic biosignals in medical diagnostics (on an example of radial artery pulse signal)

*Pokrovskaya Irina*¹*

ivp750@mail.ru

*Guchuk Vladimir*¹

polma@bk.ru

*Desova Aelita*¹

adesova@mail.ru

Dorofeyuk Alexander^{1,2}

daa2@mail.ru

*Dorofeyuk Yuliya*¹

dorofeyuk_julia@mail.ru

¹Moscow, Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences

²Moscow, National Research University “Higher School of Economics”

The methods of classification and pattern recognition for the analysis of quasi-periodic biosignals in medical diagnostics are described. It was implemented as a hardware-software complex (HSC), producing, encoding, recording in the database, visualizing, and mining in real-time of the biosignal multivariate quasi-periodic characteristics. The consideration is carried out on an example of the radial artery pulse signal processing. The complex includes: a sensor signal, device digitizing the analog signal, and the software developed by the authors on the basis of CodeGear RAD Studio. The HSC has shown its high efficiency in solving applied problems in medical diagnostics, in particular, based on the results of the analysis of rhythmic structure and form of radial artery, the problem of early diagnosis of hypertension in children and adolescents has been solved [1].

The research is performed under partial financial support of the Russian Foundation for Basic Research, grants 13-07-00229, 13-07-00992, 13-07-12201, and 14-07-00463.

- [1] Guchuk, V., Desova A., Dorofeyuk A., and Kiseleva N. 2014. Hardware and software methods for the quasi-periodic biosignals analysis (the radial artery pulse signal, as an example). *Sensors Syst.* 8:17–22.

Технология создания атласа парциальных спектров отделов головного мозга человека

*Рыкунов Станислав Дмитриевич** rykunov@psn.ru
Оплачко Екатерина Сергеевна oplachkoe@gmail.com
Полянин Андрей Георгиевич polyandroid@ya.ru
Устинин Михаил Николаевич ustinin@impb.ru

Пуцзино, Институт математических проблем биологии РАН

Предложен метод выделения парциальных спектров активности головного мозга человека по данным реконструкции функциональной структуры мозга человека на основе магнитоэнцефалографических измерений. Под парциальным спектром в данной работе понимается многоканальный спектр Фурье, порождаемый источниками, расположенными в определенной области мозга. Технология включает в себя следующие этапы: построение функциональной томограммы по данным магнитной энцефалографии; сегментация магниторезонансной томограммы головы конкретного субъекта; создание масок на основе сегментации и их наложение на функциональную томограмму. Результатом применения данной технологии являются парциальные спектры активности головного мозга человека, соответствующие определенным анатомическим структурам. Были получены парциальные спектры таламуса, мозжечка и затылочной доли мозга для 20 наборов экспериментальных данных. Пространственное разрешение полученных атласов — не менее 3 мм, частотное — не менее 1/300 Гц.

Предлагаемая технология позволит делать выводы о происхождении тех или иных деталей спектра, а также восстанавливать временной ряд энцефалограммы изучаемой области мозга [1].

Работа выполнена при частичной поддержке РФФИ, проекты 14-07-31309, 14-07-00636, 13-07-00162 и 13-07-12183, а также при поддержке Программы фундаментальных исследований Президиума РАН 43П.

- [1] *Llinás R. R., Ustinin M. N., Rykunov S. D., Boyko A. I., Sychev V. V., Walton K., Rabello G., Garcia J.* Reconstruction of human brain spontaneous activity based on frequency-pattern analysis of magnetoencephalography data // *Front. Human Neurosci.*, 2015 (in press).

Technology to build the atlas of the partial spectra of the human brain

*Rykunov Stanislav**

rykunov@psn.ru

Oplachko Ekaterina

oplachkoe@gmail.com

Polyanin Andrey

polyandroid@ya.ru

Ustinin Mikhail

ustinin@impb.ru

Pushchino, Institute of Mathematical Problems of Biology of the Russian Academy of Sciences

The method for the extraction of partial spectra of human brain activity, based on functional structure reconstruction from magnetic encephalography (MEG) measurements, is proposed. In this study, the term “partial spectrum” refers to multichannel Fourier spectrum, produced by field sources, which are located in the particular compartment of the brain. Technology consists of the following steps: computation of functional tomogram, based on MEG measurements; subject’s brain segmentation; and generation of segment’s masks and their overlapping with functional tomogram. The results of this technology implementation are the partial spectra of brain activity, corresponding to the brain anatomical structures. Partial spectra of thalamus, cerebellum, and occipital lobe were calculated for 20 experimental data sets. Spatial resolution of these atlases is 3 mm and frequency resolution is 1/300 Hz.

The proposed technology will allow to draw conclusions about origin of some spectral details. Furthermore, it will allow one to restore time series of encephalograms for particular compartment of the brain [1].

The study was partly supported by the Russian Foundation for Basic Research (grants 14-07-31309, 14-07-00636, 13-07-00162, and 13-07-12183) and by the Program 43 for Fundamental Research of the Russian Academy of Sciences.

- [1] Llinás, R. R., Ustinin M. N., Rykunov S. D., Boyko A. I., Sychev V. V., Walton K., Rabello G., and Garcia J. 2015 (in press). Reconstruction of human brain spontaneous activity based on frequency-pattern analysis of magnetoencephalography data. *Front. Human Neurosci.*

Метод частотно-временного анализа совместных измерений электроэнцефалограммы, электромиограммы и механического тремора при болезни Паркинсона

Сушкова Ольга Сергеевна^{1*} o.sushkova@mail.ru
*Габова Александра Васильевна*² agabova@yandex.ru
*Карабанов Алексей Вячеславович*³ doctor.karabanov@mail.ru
*Кершнер Иван Андреевич*⁴ ivan.kershner@gmail.com
*Обухов Константин Юрьевич*⁴ ko.arsenalfc@gmail.com
*Обухов Юрий Владимирович*¹ yuvobukhov@mail.ru

¹Москва, Институт радиотехники и электроники им. В. А. Котельникова РАН

²Москва, Институт высшей нервной деятельности и нейрофизиологии РАН

³Москва, Научный центр неврологии РАМН

⁴Москва, Московский физико-технический институт (ГУ)

Описан новый метод анализа частотно-временной динамики фоновой электрической активности мозга, с помощью которого обнаружены, по меньшей мере, три основных признака болезни Паркинсона (БП) в ранней стадии: (1) межполушарная асимметрия частотно-временных характеристик электроэнцефалограммы (ЭЭГ) в центральных отведениях моторной зоны коры головного мозга; (2) возникновение ритма ЭЭГ в этих отведениях в частотном диапазоне 4–6 Гц и его связанность с электромиограммами и механическим тремором контралатеральных конечностей при дрожательной форме БП; (3) дезорганизация доминирующего ритма, соответствующая общим представлениям о дезорганизации различных систем при БП [1].

Работа поддержана грантом РФФИ № 15-07-07846-А и Программой Президиума РАН «Фундаментальные науки — медицине».

- [1] Сушкова О. С., Габова А. В., Карабанов А. В., Кершнер И. А., Обухов К. Ю., Обухов Ю. В. Метод частотно-временного анализа совместных измерений ЭЭГ, ЭМГ и механического тремора при болезни Паркинсона // *Нелинейный мир*, 2015. № 2. Т. 13. С. 49–51. <http://www.radiotec.ru/catalog.php?cat=jr11&art=16023>.

A method of time-frequency analysis of joint electroencephalogram, electromyogram, and mechanical tremor signals in Parkinson's disease

Sushkova Olga^{1*}

`o.sushkova@mail.ru`

*Gabova Alexandra*²

`agabova@yandex.ru`

*Karabanov Alexey*³

`doctor.karabanov@mail.ru`

*Kershner Ivan*⁴

`ivan.kershner@gmail.com`

*Obukhov Konstantin*⁴

`ko.arsenalfc@gmail.com`

*Obukhov Yuri*¹

`yuvobukhov@mail.ru`

¹Moscow, Kotel'nikov Institute of Radioengineering and Electronics of the Russian Academy of Sciences

²Moscow, Institute of Higher Nervous Activity and Neurophysiology of the Russian Academy of Sciences

³Moscow, Scientific Centre of Neurology of the Russian Academy of Medical Science

⁴Moscow, Moscow Institute of Physics and Technology (State University)

A new method of analyzing time-frequency dynamics of background electrical activity of the brain is described by means of which at least three main features of Parkinson's disease (PD) in the early stage are detected: (*i*) interhemispheric asymmetry of the time-frequency characteristics at electroencephalograms (EEG) in the central leads of the motor area of the cerebral cortex; (*ii*) the emergence of EEG rhythm in these leads in the frequency range of 4–6 Hz and its relatedness to the electromyograms and mechanical tremor of the contralateral limbs in the tremor form of PD; and (*iii*) disorganization of the dominant rhythm corresponding to the general ideas about the disorganization of the various systems in Parkinson's disease [1].

This research is funded by the Russian Foundation for Basic Research, grant 15-07-07846-A and by the Program of the Presidium of the Russian Academy of Sciences "Fundamental sciences to medicine."

- [1] Sushkova O., Gabova A., Karabanov A., Kershner I., Obukhov K., and Obukhov Y. 2015. A method of time-frequency analysis of joint EEG, EMG, mechanical tremor signals in Parkinson's disease. *Nelineinyi Mir* [Nonlinear World] 13(2):49–51. <http://www.radiotec.ru/catalog.php?cat=jr11&art=16023>.

Reconstruction of human brain spontaneous activity based on frequency–pattern analysis of magnetoencephalography data

*Ustinin Mikhail**

ustinin@impb.ru

Rykunov Stanislav

rykunov@psn.ru

Boyko Anna

a.boyko@list.ru

Sychev Vyacheslav

sychyov@yahoo.com

Pushchino, Institute of Mathematical Problems of Biology of the Russian Academy of Sciences

A new method for the analysis and localization of brain activity has been developed based on multichannel magnetic field recordings. A high-resolution Fourier transform is obtained over the entire recording period, leading to a detailed multifrequency spectrum. Further analysis implements a total decomposition of the frequency components into functionally invariant entities, each having an invariant field pattern. The method, called “functional tomography,” makes it possible to find the distribution of magnetic field sources in space. Concerning the dipole phantom, the method is able to localize three simultaneously activated current dipoles with 1-millimeter precision. Spatial resolution 3 mm was attained when localizing spontaneous alpha rhythm activity in ten healthy adults, where the alpha frequency was specified for each subject individually. The method successfully provides an overall view of brain electrical activity, a detailed spectral description, and, combined with magnetic resonance imaging, the localization of sources in anatomical brain space [1].

The study was partly supported by the Russian Foundation for Basic Research (grants 13-07-00162, 13-07-12183, 14-07-00636, and 14-07-31309) and by the Program 43 for Fundamental Research of the Russian Academy of Sciences.

- [1] Llinás, R. R., Ustinin M. N., Rykunov S. D., Boyko A. I., Sychev V. V., Walton K., Rabello G., and Garcia J. 2015 (in press). Reconstruction of human brain spontaneous activity based on frequency-pattern analysis of magnetoencephalography data. *Front. Human Neurosci.*

Сравнение информационного анализа электрокардиосигналов и анализа variability сердечного ритма как методов диагностики заболеваний внутренних органов

Целых Влада Руслановна^{1*} celyh@phystech.edu

Воронцов Константин Вячеславович^{1,2} voron@forecsys.ru

*Успенский Вячеслав Максимилианович*³ medddik@yandex.ru

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Вычислительный центр им. А. А. Дородницына РАН

³Москва, Медицинский учебно-научный клинический центр им. П. В. Мандрыка

Технология информационного анализа электрокардиосигналов, предложенная В. М. Успенским для диагностики заболеваний внутренних органов по электрокардиограмме (ЭКГ), основана на кодировании знаков приращений интервалов и амплитуд R-пиков ЭКГ в символьную последовательность (кодограмму). В качестве признаков для распознавания заболеваний используются частоты коротких фрагментов кодограммы. Предшествующие исследования показали, что для каждого заболевания специфичен свой набор часто встречающихся фрагментов. Целью данной работы является повышение качества диагностики путем подбора способов кодирования и оптимизации моделей классификации. Предлагаемые методы сравниваются с анализом variability сердечного ритма (VSR), использующим только агрегированные данные об RR-интервалах. Предложены эффективные методы отбора признаков за время, линейное по длине выборки и числу признаков. Эксперименты проводились на выборке из 7541 обследования с диагнозами по 19 заболеваниям. Совместный учет variability интервалов и амплитуд повышает AUC на тестовых данных в среднем на 4%–5% по сравнению с 18 стандартными параметрами VSR-анализа. Для отдельных болезней точность диагностики достигает 98% по критерию AUC на тестовой выборке [1].

Работа поддержана грантами РФФИ 14-07-31163 и 14-07-00908.

[1] Успенский В. М., Воронцов К. В., Целых В. Р., Бунаков В. А., Ишккина Ш. Х. Технология информационного анализа электрокардиосигналов: статистическая экспертиза и пути совершенствования // Клиническая медицина, 2015 (в печати).

The comparison of the information analysis of ECG-signals and the heart rate variability analysis as techniques for the diagnostics of internal diseases

*Tselykh Vlada*¹★

celyh@phystech.edu

Vorontsov Konstantin^{1,2}

voron@forecsys.ru

*Uspenskiy Vyacheslav*³

meddik@yandex.ru

¹Moscow, Moscow Institute of Physics and Technology (State University)

²Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

³Moscow, Mandryk Medical Educational and Scientific Clinical Center

The information analysis of the ECG signals is a machine learning technique proposed by V. M. Uspenskiy to diagnose internal diseases. It is based on the encoding of the sequence of RR-intervals and R-peak amplitudes into a character string called a codegram. The frequencies of n -character fragments (called n -grams or codewords) are used as feature vector representation of each ECG-record. Previous experiments show that each disease manifests itself by a high frequency of specific codewords. Each frequent codeword describes a pattern of heart rate variability specific for the disease. The purpose of the present work is to improve the diagnostics accuracy by optimization of encoding techniques and classification models. The diagnostic algorithms are compared with the state-of-the art heart rate variability (HRV) analysis based on the descriptive statistics of the RR-intervals variability. Note that usual HRV-analysis uses only the RR-intervals and never uses the amplitudes of R-peaks. The authors propose computationally effective feature selection algorithm that finds diagnostic patterns of variability with linear time complexity both in the sample size and the number of features. For the experimental study, a dataset of 7541 ECG records with known diagnoses of 19 diseases has been used. The optimal classification model uses bigrams and a 6-character encoding. It gives the average increase of cross-validated AUC by 4%–5% if compared with HRV-analysis based on 18 standard parameters. For certain diseases, the cross-validated AUC reaches 98% [1].

The work is supported by the Russian Foundatio for Basic Research, grants 14-07-31163 and 14-07-00908.

[1] Uspenskiy, V., Vorontsov K., Tselykh V., Bunakov V., and Ishkina Sh. 2015 (in press). Informational analysis of ECG signals: Statistical verification and ways of improvement. *Klinicheskaya Meditsina*.

Параллельные тексты в задаче дешифровки древнерусских знаменных песнопений

Бахмутова Ирина Владимировна bakh@math.nsc.ru
Гусев Владимир Дмитриевич gusev@math.nsc.ru
Мирошниченко Любовь Александровна luba@math.nsc.ru
Титкова Татьяна Николаевна titkova@math.nsc.ru

Новосибирск, Институт математики им. С. Л. Соболева СО РАН

Древнерусские церковные песнопения XII–XVII вв. представлены в знаменной форме записи. Проблема перевода их в нотолинейную форму носит дешифровочный характер и в общем случае (при отсутствии у знамен помет, поясняющих их певческое значение) остается открытой. Количество непрочитанных певческих книг исчисляется сотнями. Трудности дешифровки связаны с многозначностью соответствия «знамя – нота».

Авторы развивают новый подход к решению этой проблемы, основанный на использовании двознаменников — певческих книг конца XVII — начала XVIII вв., записанных в виде трех параллельных (синхронизованных между собой) текстов: знаменного, нотолинейного и стихотворного (старославянского). Акцент делается на выявление в текстах двознаменников повторяющихся цепочек знамен, интерпретируемых однозначно (инварианты) или «почти однозначно» (квазиинварианты).

На обучающем материале построены электронные словари инвариантов и квазиинвариантов. Разработан алгоритм дешифровки песнопений с использованием этих словарей. Эксперименты на контрольном материале показали, что даже без апелляции к структурной организации знаменного распева удается правильно дешифровать порядка 60%–70% знаменного текста.

Особенностями подхода являются: опора на многожанровые двознаменники периода наивысшего расцвета знаменного пения и ориентация на общий случай беспометной нотации [1].

Работа поддержана грантом РФФИ № 13-07-00400.

- [1] *Бахмутова И. В., Гусев В. Д., Мирошниченко Л. А., Титкова Т. Н.* Параллельные тексты в задаче дешифровки древнерусских знаменных песнопений // Машинное обучение и анализ данных, 2015 (в печати). Т. 1. №. 13. jmla.org/papers/doc/2015/JMLDA2015no13.pdf.

Parallel texts in the problem of deciphering of ancient Russian Chant

Bakhmutova Irina

bakh@math.nsc.ru

Gusev Vladimir

gusev@math.nsc.ru

*Miroshnichenko Liubov**

luba@math.nsc.ru

Titkova Tatiana

titkova@math.nsc.ru

Novosibirsk, Sobolev Institute of Mathematics of the Siberian Branch
of the Russian Academy of Sciences

The ancient Russian chants of the XII–XVII centuries are presented in the neume writing form. The problem of chant translation into note writing is of deciphering character and, in the general case (chants without marks that explain their singing value), is not yet solved. The number of “unreadable” ancient hymnals runs into the hundreds. The difficulties of deciphering are connected with the polysemy of the correspondence “neume–note.” The known examples of deciphering are few in number, made manually, and refer to the certain hymn.

The authors develop a new approach to the solution of this problem using the dvoeznamenniks of the XVII–XVIII centuries where the chants are written in three (synchronized with each other) parallel texts: in neumes, in notes, and in old Slavonic verses. The emphasis is made on revealing in texts of dvoeznamenniks repeating chains of neumes that are interpreted either unambiguously (invariants) or with admissible deviations (quasi-invariants)

On the basis of learning material, the electronic dictionaries of invariants and quasi-invariants have been constructed. The algorithm of deciphering of neumatic notation using these dictionaries was developed. The experiments on the control material have shown that these dictionaries provide the deciphering of 60%–70% of text.

The main features of the present approach are: use of dvoeznamenniks in different genres and orientation, in general, toward the chants without special marks from the XVI–XVII centuries [1].

This research is funded by Russian Foundation for Basic Research, grant 13-07-00400.

- [1] Bakhmutova, I. V., Gusev V. D., Miroshnichenko L. A., Titkova T. N. 2015. Parallel texts in the problem of deciphering of ancient Russian Chant. *J. Mach. Learn. Data Anal.* 1(13). jmla.org/papers/doc/2015/JMLDA2015no13.pdf.

Регуляризация тематических моделей в библиотеке с открытым кодом BigARTM

Воронцов Константин Вячеславович^{1,2,3*} voron@forecsys.ru

*Фрей Александр Ильич*¹ oleksandr.frei@gmail.com

*Апишев Мурат Азаматович*³ great-mel@yandex.ru

*Дойков Никита Владимирович*³ nikitad101@gmail.com

*Суворова Марина Алексеевна*³ m.dudarenko@gmail.com

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Вычислительный центр им. А. А. Дородницына РАН

³Москва, Московский государственный университет
им. М. В. Ломоносова

Аддитивная регуляризация тематических моделей (ARTM) позволяет выявлять латентную тематику коллекций текстовых документов, решая некорректно поставленную задачу стохастического матричного разложения. В библиотеке BigARTM, доступной по адресу bigartm.org, реализован однопроходный параллельный EM-алгоритм, превосходящий другие библиотеки по времени выполнения и по гибкости конструирования тематических моделей. BigARTM позволяет, комбинируя регуляризаторы, получать тематические модели с заданными свойствами.

Рассматриваются регуляризаторы для повышения интерпретируемости тем путем разреживания, сглаживания и декоррелирования; устранения дублирующих и линейно зависимых тем; построения мультимодальных моделей по метаданным о времени, авторах, пользователях, ссылках и других связанных с документами объектах; построения мультязычных тематических моделей для кросс-язычного поиска; разделения тем на событийные и перманентные; выявления иерархических связей между темами; выделения тематических ключевых фраз; тематической сегментации текста. Рассматривается обобщенная гиперграфовая тематическая модель для описания социальных сетей, рекламных сетей и рекомендательных систем. Рассматриваются приложения тематического моделирования [1].

Работа поддержана РФФИ, гранты №№ 14-07-00847, 14-07-00908 и 14-07-31176.

[1] Воронцов К. В., Фрей А. И., Апишев М. А. и др. BigARTM: библиотека с открытым кодом для тематического моделирования больших текстовых коллекций // DAMDID, 2015.

Regularization of topic models in BigARTM open source library

Vorontsov Konstantin^{1,2,3*}

voron@forecsys.ru

*Frei Olexandr*¹

oleksandr.frei@gmail.com

*Apishev Murat*³

great-mel@yandex.ru

*Doikov Nikita*³

nikitad101@gmail.com

*Suvorova Marina*³

m.dudarenko@gmail.com

¹Moscow, Moscow Institute of Physics and Technology (State University)

²Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

³Moscow, Lomonosov Moscow State University

Additive Regularization of Topic Models (ARTM) reveals a hidden thematic structure of a text collection by solving an ill-posed problem of stochastic matrix factorization. The online parallel EM-algorithm for ARTM was implemented in the open source library BigARTM available at bigartm.org. BigARTM outperforms concurrent topic modeling tools by both running time and flexibility of model construction. BigARTM allows to produce topic models with the desired properties by combining regularizers.

The authors consider regularizers for topic sparsing, smoothing, and decorrelation that augments the interpretability of topics; topic selection that eliminates duplicate and linearly dependent topics; multimodal likelihood maximization that takes into account document metadata such as time, authors, users, links, etc.; multilingual topic modeling for cross-lingual information search; temporal modeling that separates event topics and permanent topics; hierarchical modeling that induces relationships between topics and their subtopics; and segmentation modeling that identifies thematically homogeneous fragments in a coherent text. Then, the authors consider a hypergraph generalization of ARTM for joint modeling of text and transactional data in social media, advertising networks and recommender systems. Finally, experimental results are reported on several applications of ARTM [1].

The work was supported by the Russian Foundation for Basic Research, grants 14-07-00847, 14-07-00908, and 14-07-31176.

- [1] Vorontsov, K., Frei O., Apishev M., *et al.* 2015. BigARTM: Open source library for topic modeling of large text collections. *DAMDID*.

Построение иерархических тематических моделей крупных конференций

Кузьмин Арсентий Александрович^{1*}

arsentii.kuzmin@gmail.com

*Адуенко Александр Александрович*¹

aduenko1@gmail.com

*Стрижов Вадим Викторович*²

strijov@ccas.ru

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Вычислительный центр им. А. А. Дородницына РАН

Решается проблема построения и верификации иерархических тематических моделей крупных конференций. Предполагается, что тематическая модель конференции изменяется из года в год незначительно, что позволяет фиксировать ее иерархическую структуру и классифицировать новые аннотации докладов.

Для классификации новых аннотаций вводится оператор релевантности, ранжирующий кластеры нижнего уровня тематической модели в порядке убывания сходства. Существующие методы оценки сходства между документом и кластером нижнего уровня иерархии не являются устойчивыми в силу небольшого размера кластеров. Добавление нового документа в кластер приводит к значительному изменению его терминологического состава и, как следствие, сходства между данным кластером и уже находящимися в нем документами. Предлагается использовать иерархическую взвешенную функцию сходства документа и кластера, учитывающую сходство документа со всеми его родительскими кластерами. Весами в данной функции сходства являются значения важности терминов из словаря коллекции, которые выражаются через их энтропию относительно экспертной кластеризации на заданном уровне иерархии [1].

Для оценки результатов вводятся два критерия качества оператора релевантности. Для проверки предлагаемых алгоритмов строится тематическая модель конференции EURO 2010 по экспертным моделям конференций EURO 2012 и EURO 2013.

Работа поддержана грантом РФФИ № 14-07-31264.

- [1] *Кузьмин А. А., Адуенко А. А., Стрижов А. А.* Тематическая классификация тезисов крупной конференции с использованием экспертной модели // Информационные технологии, 2014. Т. 6. С. 22–26.

Hierarchical thematic modeling of major conference proceedings

Kuzmin Arsentii^{1*}

arsentii.kuzmin@gmail.com

*Aduenko Alexander*¹

aduenko1@gmail.com

*Strijov Vadim*²

strijov@ccas.ru

¹Moscow, Moscow Institute of Physics and Technology (State University)

²Moscow, Dorodnicyn Computing Center of the Russian Academy of Sciences

The aim of this study is to construct and verify a hierarchical thematic model of a major conference. Only a few themes of the conference change from one year to another. Here, previous conferences hierarchical structures have been used to classify new abstracts of the upcoming conference.

A relevance operator is suggested to classify new unmarked abstracts (documents). It sorts the clusters of the lowest level of hierarchy according to similarity with the new document in the descending order. There are several ways to calculate a similarity of a document and a cluster of the lowest level of hierarchy using the term frequencies of the document and the cluster. But adding a new document to such a cluster changes its term frequencies significantly. It leads to significant changes of similarity values of this cluster and its documents. A weighted hierarchical similarity function is proposed to avoid this problem. It considers similarities of document and all parent clusters of the lowest level cluster. Weights in this function are the importance values of the terms from the collection dictionary. Entropy-based approach is used to estimate these weights [1].

The proposed relevance operator is compared with its baseline pseudorandom version using two proposed quality criterions. The suggested method constructs a thematic model for the EURO 2010 conference using the expert models for the conferences EURO 2012 and EURO 2013.

This research is funded by the Russian Foundation for Basic Research, grant 14-07-31264.

- [1] Kuzmin, A., Aduenko A., and Strijov V. 2014. Thematic classification using expert model for major conference abstracts . *Informational Technologies* 6:22–26.

Мера TF-IDF и формирование единиц представления знаний для открытых тестов

*Михайлов Дмитрий Владимирович**

mdv74@list.ru

Козлов Александр Павлович

caleo@yandex.ru

Емельянов Геннадий Мартинович Gennady.Emelyanov@novsu.ru

Великий Новгород, Новгородский государственный университет

Разработка открытых тестов предполагает накопление и систематизацию экспертных знаний, исходно представляемых текстами естественного языка (ЕЯ). Источником знаний здесь будут публикации отечественных и зарубежных научных школ в виде монографий, обзорных статей, трудов конференций и т. п. В работе предложен метод поиска в текстовом корпусе фраз, максимально близких исходной по описываемому фрагменту знания и формам его выражения в заданном ЕЯ. В основе метода лежит разбиение слов исходной фразы на классы по значению меры TF-IDF (term frequency – inverse document frequency) относительно документов корпуса для выделения общей лексики, обеспечивающей синонимические перифразы, а также слов-терминов. Конечной целью является поиск наиболее рационального плана передачи смысла средствами заданного ЕЯ между экспертами и обучаемыми. Одним из базовых здесь служит предположение о том, что при наличии перифраз задающая их общая лексика имеет значения TF и IDF, близкие средним по исходной фразе. Сортируя по убыванию TF-IDF, наиболее значимыми выделяют три класса слов: близких максимальному, минимальному и среднему значениям указанной меры. По результатам кластеризации слов исходной фразы отбираются документы корпуса, относительно которых найдено как можно больше кластеров при минимуме разности наибольшего и наименьшего диаметров и максимально возможном числе слов кластера. Результирующие фразы из выделенных документов выбираются по максимуму слов в трех наиболее значимых классах по TF-IDF [1].

Работа поддержана РФФИ (проект № 13-01-00055) и Минобрнауки РФ (базовая часть госзадания).

[1] Михайлов Д. В., Козлов А. П., Емельянов Г. М. Выделение знаний и языковых форм их выражения на множестве тематических текстов: подход на основе меры TF-IDF // Компьютерная оптика, 2015, Т. 39, №3, С. 429-438. www.computeroptics.smr.ru/KO/PDF/K039-3/390319.pdf.

TF-IDF metrics and formation of units for knowledge representation in open tests

*Mikhaylov Dmitry**

mdv74@list.ru

Kozlov Aleksandr

caleo@yandex.ru

Emelyanov Gennady

Gennady.Emelyanov@novsu.ru

Veliky Novgorod, Yaroslav-the-Wise Novgorod State University

Development of open form test assignments requires the accumulating and ordering an expert knowledge initially represented by natural-language (NL) texts. The most natural knowledge source here will be the scientific papers of highest rank scholars in appropriated topical area. In this paper, the method to select in text corpora the phrases closest to initial concerning the described fragment of reality or expressional forms for expert knowledge in given NL is suggested. This method is based on clustering the vocabulary of initial phrase by TF-IDF (term frequency – inverse document frequency) metrics relatively to corpora documents with division between terms and general vocabulary by means of which synonymic paraphrases can be formed. The main practical goal here is to find the most rational plan to transfer the meaning between experts and testees using the given NL. One of the base assumptions here is that when paraphrases are present, the general vocabulary which specifies them has the values of both TF and IDF closest to average for initial phrase. Sorting the words by descending values of TF-IDF, as the most significant three word classes are selected: closest to maximal, minimal, and average value of the given measure. The results of word clustering are the basis of selection the corpora documents relatively to which the largest number of clusters were found at minimal difference between the largest and the least values of cluster width and maximum of words in each cluster. The resultant phrases from selected documents are chosen according to the criteria of maximum of words related to three most significant clusters by TF-IDF [1].

This research is funded by the Russian Foundation for Basic Research, grant 13-01-00055, and the Ministry of Education and Science of Russia (the basic part of the state task).

- [1] Mikhaylov, D., Kozlov A., and Emelyanov G. 2015. An approach based on TF-IDF metrics to extract the knowledge and relevant linguistic means on subject-oriented text sets. *Comput. Opt.* 39(3):429–438. www.computeroptics.smr.ru/KO/PDF/K039-3/390319.pdf.

Энтропийный регуляризатор отбора тем в вероятностных тематических моделях

Плавин Александр Викторович^{1*} alexander@plav.in
*Потапенко Анна Александровна*² anya_potapenko@mail.ru
Воронцов Константин Вячеславович^{1,2,3} voron@forecsys.ru

¹Москва, Московский физико-технический институт (ГУ)

²Москва, НИУ «Высшая школа экономики»

³Москва, Вычислительный центр им. А. А. Дородницына РАН

В задачах машинного обучения важным этапом решения является определение структурного параметра модели. Для регрессии и классификации — это число используемых признаков, для кластеризации — число кластеров, для матричного разложения — промежуточная размерность матриц или число главных компонент. В вероятностном тематическом моделировании коллекций текстовых документов структурным параметром является число тем. От его выбора зависят такие характеристики качества тем, как интерпретируемость, согласованность, различность.

Предлагается энтропийный регуляризатор, который удаляет наименее значимые темы постепенно в ходе итераций. Он позволяет верно определять заранее известное число тем для синтетических коллекций. Однако на реальных данных четко определить число тем, как правило, не удастся. Стандартной моделью для определения числа тем считается иерархический процесс Дирихле. Эксперименты показывают, что в обеих моделях результат зависит от произвольно выбираемого значения гиперпараметра. Тем не менее, энтропийный регуляризатор имеет ряд преимуществ: он в десятки раз быстрее, свободно сочетается с другими регуляризаторами и определяет число тем более устойчиво как в одном запуске, так и при нескольких запусках из различных начальных приближений. Кроме того, он очищает модель от линейно зависящих тем и дублирующих вложенных подтем, что способствует повышению интерпретируемости модели [1].

Работа поддержана грантом РФФИ № 14-07-00908.

- [1] Vorontsov K. V., Potapenko A. A., Plavin A. V. Additive regularization of topic models for topic selection and sparse factorization // Statistical learning and data sciences / Eds. A. Gammerman, V. Vovle, H. Papadopoulos. — Lecture notes in artificial intelligence ser. — Switzerland: Springer International Publishing, 2015. Vol. 9047. P. 193–202.

Topic selection in probabilistic topics models with entropy regularization

*Plavin Alexander*¹*

alexander@plav.in

*Potapenko Anna*²

anya.potapenko@mail.ru

Vorontsov Konstantin^{1,2,3}

voron@forecsys.ru

¹Moscow, Moscow Institute of Physics and Technology (State University)

²Moscow, National Research University “Higher School of Economics”

³Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

Determining the model structure is an important problem in machine learning tasks. In regression and classification, the structural parameter stands for the number of selected features; in matrix factorization, it corresponds to the inner dimension or the number of principal components. In this work, the probabilistic topic models of text collections, where the number of topics is the parameter to be chosen, are considered. This is an open problem essential for obtaining well interpretable, coherent, and diverse topics.

The authors propose the entropy regularization criteria that removes less significant topics at each iteration for additively regularized topic model (ARTM). It is shown that the proposed approach determines the correct number of topics for synthetic data. For real data with unknown true number of topics, the entropy regularizer is compared with the state-of-the-art Hierarchical Dirichlet Processes (HDP) topic model. It is shown that the resulting number of topics depends on an arbitrarily assigned hyperparameter in both methods. Nevertheless, the present approach has several advantages: it is much faster, it can be freely combined with other regularization criteria, and it gives a more stable estimate for the number of topics both from run to run and within one run. Furthermore, it is shown theoretically and experimentally that the entropy regularizer eliminates excessive topics: linear combinations of other topics, duplicate topics, and nested subtopics. Thus, the topic selection based on the entropy regularization increases the interpretability of the topic model [1].

The work was supported by the Russian Foundation for Basic Research, grant 14-07-00908.

- [1] Vorontsov, K. V., Potapenko A. A., and Plavin A. V. 2015. Additive regularization of topic models for topic selection and sparse factorization. *Statistical learning and data sciences*. Eds. A. Gammerman, V. Vovle, and H. Papadopoulos. Lecture notes in artificial intelligence ser. Switzerland: Springer International Publishing. 9047:193–202.

Иерархическая аддитивно регуляризованная тематическая модель научной конференции

Чиркова Надежда Александровна^{1*} nadiinchi@gmail.com

*Айсина Роза Мунеровна*¹ rose.aysina@gmail.com

Воронцов Константин Вячеславович^{1,2} vokov@forecsys.ru

¹Москва, Московский государственный университет
им. М. В. Ломоносова

²Москва, Вычислительный центр им. А. А. Дородницына РАН

Вероятностная тематическая модель текстовой коллекции описывает каждый документ дискретным распределением на множестве тем, а каждую тему — дискретным распределением на множестве терминов (слов или фраз). Иерархическая тематическая модель — это многодольный ориентированный граф тем, в котором каждая дочерняя тема является специализацией родительской темы. Тематические иерархии используются для категоризации текстов, систематизации и визуализации больших коллекций. Задача автоматического построения и оценивания тематических иерархий считается открытой научной проблемой.

В данной работе предлагается нисходящий метод построения иерархии, основанный на *аддитивной регуляризации тематической модели* (АРТМ). Предлагается несколько вариантов иерархического регуляризатора, связывающего родительские и дочерние темы между уровнями и допускающий множественное наследование тем. Дополнительные регуляризаторы разреживания, сглаживания и декоррелирования улучшают интерпретируемость тем и позволяют выделять общую лексику каждой внутренней вершины иерархии. Иерархия строится автоматически, уровень за уровнем. Эксперт задает число уровней, стартовое число тем на каждом уровне и коэффициенты регуляризации.

Эксперименты на коллекции 850 статей научных конференций ММРО и ИОИ за последние 6 лет показали, что интерпретируемость иерархии существенно улучшается при использовании мультиграммных терминов. Веб-интерфейс иерархической модели доступен по адресу <http://explore-mmro.ru>.

Работа поддержана РФФИ, грант № 14-07-31176.

- [1] *Чиркова Н. А.* Иерархическая аддитивно регуляризованная тематическая модель коллекции научных текстов // Машинное обучение и анализ данных, 2015 (в печати).

Hierarchical additively regularized topic model of a scientific conference

Chirkova Nadia^{1*}

nadiinchi@gmail.com

*Aysina Roza*¹

rose.aysina@gmail.com

Vorontsov Konstantin^{1,2}

vokov@forecsys.ru

¹Moscow, Lomonosov Moscow State University

²Moscow, Dorodnicyn Computing Centre of the Russian Academy of Sciences

Probabilistic topic model describes a text collection by document-wise probability distributions over topics and topic-wise probability distributions over terms (words or key phrases). Hierarchical topic model organizes topics as an oriented multipartite graph where each child topic describes a specialized part of the parent topic terminology. Topic hierarchy is used for categorization of texts, navigation, visualization, and exploratory search. Fully automatic construction and evaluation of topic hierarchies of large text collections are considered to be an open question.

In this work, the authors propose a top-down approach to building topic hierarchies, based on *additive regularization of topic models* (ARTM). Hierarchy is built level by level. Several variants of a hierarchical regularizer are proposed to connect topics between levels and to determine the number of topics at each level. This hierarchical regularizer supports multiple parents per topic. Sparsing, smoothing, and decorrelation regularizers improve topic interpretability and separate a background parent topic terminology from a more specific child topic terminology for each inner node of the hierarchy. The hierarchy is built automatically; an expert specifies only number of levels, starting number of topics on each level, and regularization coefficients [1].

Experiments on 850 papers collection of MMRP and IIP proceedings from 2007 to 2013 show that the hierarchy interpretability significantly improves when multigram terms are used. The MMRO-IIP hierarchy is available at <http://explore-mmro.ru>.

The work was supported by the Russian Foundation for Basic Research, grant 14-07-31176.

[1] *Chirkova, N.* 2015 (in press). Hierarchical additively regularized topic model of scientific document collection. *J. Mach. Learn. Data Anal.*

Эволюция подходов к модерации контента для сайта объявлений: от рекомендаций до автоматизации

Гуз Иван Сергеевич

`iguz@avito.ru`

*Лексин Василий Алексеевич**

`vleksin@avito.ru`

Трофимов Михаил Игоревич

`mtrofimov@avito.ru`

Фенстер Александра Михайловна

`afenster@avito.ru`

Москва, Avito.ru

По мере роста популярности сайтов объявлений следить за качеством их контента становится все сложнее. Чем больше покупателей, тем привлекательнее становится ресурс для злоумышленников, желающих разместить запрещенный контент, и для продавцов, стремящихся продвинуть свой товар путем создания многочисленных копий, описание которых различается по написанию, но имеет одинаковый смысл. В определенный момент продавцы начинают писать собственных роботов, которые размещают дубликаты или запрещенные товары автоматически, используя техники морфинга текстов и подключая различные краудсорсинговые платформы для обхода любой капчи. Просматривать весь создаваемый контент с помощью наращивания команды модераторов становится нереалистичным, и требуются принципиально новые подходы для решения этой задачи. Рассматривается комплексный подход, основанный на методах машинного обучения, в котором алгоритмы последовательно обучаются на действиях экспертов, а также практический подход к внедрению этих алгоритмов. Изначально решения алгоритмов выдаются людям в виде рекомендаций с некоторой степенью уверенности, что позволяет оценить их качество и получить обратную связь. Затем, по мере подтверждения корректности, решения алгоритмов, в которых они достаточно сильно уверены, начинают исполняться автоматически без участия людей. Таким образом происходит процесс эволюции всей системы [1].

- [1] *Гуз И. С., Лексин В. А., Трофимов М. И., Фенстер А. М.* Эволюция подходов к модерации контента для сайта объявлений: от рекомендаций до автоматизации // CrowdML — ICML'15 Workshop on Crowdsourcing and Machine Learning, 2015. crowdml.cc/icml2015/papers/CrowdML-Paper5.pdf.

Evolution of content moderation approaches for online classifieds: From action recommendations to automation

Guz Ivan

iguz@avito.ru

*Leksin Vasily**

vleksin@avito.ru

Trofimov Mikhail

mtrofimov@avito.ru

Fenster Aleksandra

afenster@avito.ru

Moscow, Avito.ru

As online classified advertising platforms become more and more popular, it becomes harder to ensure quality of the available content. The more buyers they attract, the more attractive site becomes also for swindlers to upload prohibited content and for sellers who are trying to promote their ads by creating multiple copies of each ad. Those duplicate ads have different descriptions and might have different images but their meaning is the same. At certain time, sellers start to create automated tools for massive uploading and generation of ad duplicates. Text morphing algorithms are used to change visual ad description and anticaptcha crowdsourcing platforms are used to bypass any captcha. Validating all that inflow by scaling moderation team is becoming unrealistic and new approaches are required to address this challenge. Here, complex approach is proposed that is based on machine learning methods. In this approach, the algorithms are trained on actions of humans. At the beginning, the algorithms only provide certain recommendations with corresponding confidence levels which allows to estimate and get feedback on their accuracy. When their predictive accuracy becomes verified, the algorithms are allowed to start taking actions on their own in situations when their confidence is high. At the end, human moderation focuses only on the most complex verification cases while the majority of general cases are moderated automatically. This is how this human computation system evolves over time [1].

- [1] Guz, i., Leksin V., Trofimov M., and Fenster A. 2015. Evolution of content moderation approaches for online classifieds: From action recommendations to automation. *CrowdML — ICML'15 Workshop on Crowdsourcing and Machine Learning*. crowdml.cc/icml2015/papers/CrowdML-Paper5.pdf.

Структурно-классификационные методы интеллектуального анализа данных для слабо формализованных систем управления

Дорофеев Александр Александрович^{1,2*} daa2@mail.ru

*Дорофеев Юлия Александровна*¹ dorofeyuk_julia@mail.ru

*Покровская Ирина Вячеславовна*¹ ivp750@mail.ru

*Чернявский Александр Леонидович*¹ achern@ipu.ru

¹Москва, Институт проблем управления РАН

²Москва, Национальный исследовательский университет «Высшая школа экономики»

Рассмотрены особенности реализации и практического применения методов и алгоритмов структурно-классификационного анализа данных при анализе и синтезе слабо формализованных систем управления, в том числе рассмотрены методы: структуризации параметров; классификации (кластеризации) объектов; динамического структурно-классификационного анализа временных рядов; кусочной аппроксимации сложных, нелинейных динамических моделей исследуемого объекта. Эти особенности использовались при создании программно-алгоритмического комплекса (ПАК) с интеллектуальным интерфейсом, ориентированного на пользователя-предметника [1]. При разработке этого комплекса особое внимание уделялось задачам, в которых исследуемые объекты имеют явно выраженную территориальную структуру. Разработанный ПАК использовался при решении широкого класса прикладных задач, в том числе при исследовании медико-биологических, социально-экономических и финансовых систем, в задачах управления региональным здравоохранением и отдельными медицинскими учреждениями, региональными пассажирскими автоперевозками и др.

Работа выполнена при частичной финансовой поддержке РФФИ, гранты №№ 15-07-06713, 14-07-00463, 13-07-00992 и 13-07-12201.

- [1] *Киселева Н. Е., Дорофеев А. А., Дорофеев Ю. А., Покровская И. В.* Программно-алгоритмический комплекс интеллектуального анализа слабо формализованных данных // Датчики и системы, 2014. № 6. С. 48–53.

Structural classification data mining techniques for weakly formalized control systems

Dorofeyuk Alexander^{1,2,*}

daa2@mail.ru

*Dorofeyuk Yuliya*¹

dorofeyuk_julia@mail.ru

*Pokrovskaya Irina*¹

ivp750@mail.ru

*Chernyavskiy Alexander*¹

achern@ipu.ru

¹Moscow, Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences

²Moscow, National Research University "Higher School of Economics"

The paper discusses the features of the structural classification data analysis methods and algorithms implementation and practical application for the analysis and synthesis of weakly formalized management systems including the methods: of structuring parameters; of the objects classification (clustering); of the time series dynamic structural classification analysis; and of the complex nonlinear dynamic object models piecewise approximation. These features were used to create software-algorithmic complex (SAC) with smart interface, user-centered of a subject [1]. In developing this complex, special attention was paid to problems in which the studied objects have a pronounced territorial structure. Developed SAC was used to solve a wide class of applied problems, including the study of biomedical, social, economic and financial systems, in the management problems of the regional health-care systems and specific medical institutions, regional passenger transport, etc.

The research is performed under partial financial support of the Russian Foundation for Basic Research, grants: 15-07-06713, 14-07-00463, 13-07-00992, and 13-07-12201.

- [1] Kiseleva, N., Dorofeyuk A., Dorofeyuk Y., and Pokrovskaya I. 2014. Software-algorithmic complex for mining weakly formalized data. *Sensors Syst.* 6:48–53.

Модели на основе случайных лесов в задаче мониторинга состояний электроэнергетических систем

Жуков Алексей В.^{1,2*} zhukovalex13@gmail.com
Сидоров Денис Н.^{1,2} sidorovdn@mail.ru
*Томин Никита В.*¹ tomin@isem.sei.irk.ru
*Курбацкий Виктор Г.*¹ kurbatsky@isem.sei.irk.ru
*Панасецкий Даниил А.*¹ panasetsky@gmail.com
*Спиряев Вадим А.*¹ eldin@mail.ru

¹Иркутск, Институт систем энергетики им. Л. А. Мелентьева СО РАН

²Иркутск, Иркутский государственный университет

Развитие и совершенствование электроэнергетических систем (ЭЭС) и появление альтернативных источников энергии, в первую очередь ветроэнергетических установок (ВЭУ) и солнечных станций, существенно усложняют задачу мониторинга состояний энергосистем для предупреждения аварийных ситуаций. Появление цифрового оборудования дает возможность в реальном времени получать текущие значения параметров энергосистем, однако обработка таких объемов данных крайне трудоемка. В связи с этим возникает необходимость в применении современных методов анализа данных.

В данной работе рассматриваются вопросы применения моделей, основанных на случайных лесах для отбора признаков и предсказания состояний ЭЭС для решения задачи мониторинга текущего режима.

Также в работе предлагается новый подход, использующий статические и динамические модели на основе случайных лесов для решения двух практических задач — предсказание порывов ветра в ЭЭС, содержащих ВЭУ, и распознавание предаварийных состояний в ЭЭС. Приводятся результаты тестирования полученных моделей на модельных и реальных данных [1].

Работа поддержана грантом РФФИ № 14-19-00054.

- [1] *Tomin N., Zhukov A., Sidorov D., Kurbatsky V., Panasetsky D., Spiryayev V.* Random forest based model for preventing large-scale emergencies in power systems // *Int. J. Artificial Intelligence*, 2015. Vol. 13. No. 1. P. 211–228.

Random forests based models for emergency state monitoring in power systems

Zhukov Aleksi^{1,2*}

zhukovalex13@gmail.com

Sidorov Denis^{1,2}

sidorovdn@mail.ru

*Tomin Nikita*¹

tomin@isem.sei.irk.ru

*Kurbatsky Viktor*¹

kurbatsky@isem.sei.irk.ru

*Panasetsky Daniil*¹

panasetsky@gmail.com

*Spiryaev Vadim*¹

eldin@mail.ru

¹Irkutsk, Energy Systems Institute of the Siberian Branch of the Russian Academy of Sciences

²Irkutsk, Irkutsk State University

In recent years, rapid growth of power systems and wide development of renewable energy sources make the problem of power system state monitoring especially relevant. Introduction of new digital equipment allows to obtain many power system parameters in real time. However, processing of such amount of data is difficult and requires intellectual approach to data analysis.

This paper deals with the application of models based on random forests for feature selection, prediction, and classification problems in the multidimensional time series to detect alarm states in electric power systems.

Also, the authors propose a new approach using static and dynamic models based on random forests to solve two practical problems — the prediction of wind ramps in wind power system and detection of preemergency states in electric power systems. The proposed models were tested on artificial and real datasets [1].

This research is funded by the Russian Science Foundation, grant 14-19-00054.

- [1] *Tomin, N., Zhukov A., Sidorov D., Kurbatsky V., Panasetsky D., and Spiryaev V.* 2015. Random forest based model for preventing large-scale emergencies in power systems. *Int. J. Artificial Intelligence* 13(1):211–228.

Метод разработки системы многомодального распознавания казахской речи

Мамырбаев Оркен^{1*}

morkenj@mail.ru

*Калимолдаев Максат*¹

mmk@ipic.kz

*Амиргалиев Едилхан*¹

amir_ed@mail.ru

*Мусабаев Рустам*¹

rmusab@gmail.com

*Алимхан Кейлан*²

keylan@live.jp

¹Алматы, Институт информационных и вычислительных технологий
КН МОН РК

²Tokyo, Tokyo Denky University

Исследования, посвященные распознаванию речи, лица, положения человека в окружающем пространстве, ведутся уже более полувека. Многомодальные методы распознавания речи обрабатывают данные, полученные из двух и более каналов ввода информации, например, как речь и движения губ.

Созданная авторами система рассматривает двухмодальный (речь и губы) метод распознавания. По первому каналу поступает речевой сигнал от микрофона, а из второго канала поступает сигнал от видео камеры Kinect, описывающий движение губ. Разработанная многомодальная система, получая информацию из двух каналов, совместно обрабатывая их, позволяет более качественно распознавать речь. В результате установлено, что качество распознавания речи на основе двухмодального подхода выше, чем качество, полученное от отдельно взятой речи, учитывающейся в системах [1].

В данной работе исследуется метод двухмодального распознавания речи, основанный на анализе речевого сигнала и изображения лица говорящего. По результатам теоретических оценок и проведенных экспериментов установлено, что качество распознавания речи в комбинации с менее точным методом распознавания, таким как распознавание по изображению лица говорящего, повышается.

- [1] *Kalimoldayev M. N., Mamyrbayev O. J., Mussabayev R. R., Orazbekov J. N.* Segmentation and speech signal processing algorithm using the average frequency level crossing // Prob. Inform., 2014. Vol. 1. No. 22. P. 73–82.

Development method of multimodal recognition of the Kazakh speech

Mamyrbayev Orken^{1*}

morkenj@mail.ru

*Kalimoldaev Maksat*¹

mnk@ipic.kz

*Amirgaliyev Edilhan*¹

amir_ed@mail.ru

*Musabayev Rustam*¹

rmusab@gmail.com

*Alimhan Keylan*²

keylan@live.jp

¹Almaty, Institute of Information and Computational Technologies SC
MES of RK

²Tokyo, Tokyo Denky University

For more than half a century, the researches on recognition of speech, face, and human position in the surrounding area have been conducted. Multimodal methods of speech recognition process data obtained from two or more input channels, for instance, voice and lip movement.

The developed system studies bimodal (speech and lip) recognition method. Speech signal from the microphone comes through the first channel and the second channel provides signal received from the video camera Kinect, which describes the movement of the lips. This multimodal system provides higher quality of speech recognition by receiving information from two channels and processing it together. As a result, it was found that the quality of speech recognition based on the bimodal approach is higher than the one obtained from a single speech, shown in systems [1].

The method of bimodal speech recognition, which is based on the analysis of the speech signal and facial image of the speaker, is investigated. Recognition by the speech signal has relatively high, but insufficient level of recognition accuracy. The results of theoretical estimates and experiments revealed that speech recognition combined with a less accurate method of recognition, like the facial recognition, can improve the accuracy of the speech recognition.

- [1] Kalimoldayev, M.N., Mamyrbayev O.J., Mussabayev R.R., and Orazbekov J.N. 2014. Segmentation and speech signal processing algorithm using the average frequency level crossing. *Prob. Inform.* 1(22):73–82.

Идентификация имитационных моделей транспортных потоков с помощью разнородных источников прецедентной информации

*Петров Григорий Евгеньевич*¹

greekon@gmail.com

Чехович Юрий Викторович^{2*}

chehovich@forecsys.ru

¹Москва, Московский государственный университет
им. М. В. Ломоносова

²Москва, Вычислительный центр им. А. А. Дородницына РАН

Рассматривается задача вычисления плотности транспортных потоков с использованием данных из разнородных источников: транспортные детекторы и GPS-трекеры. Строится имитационная модель, позволяющая изучить способы и определить границы данных и качества данных, необходимых для восстановления характеристик транспортного потока, и предлагается схема эксперимента возможности комплексирования данных. Ключевую роль играют модели восстановления плотности транспортного потока по его скорости: модели Танака, Гриндшилдса и Гринберга.

На основе вычислительных экспериментов получены границы доли транспортных средств и точности определения местоположения для определения параметров транспортного потока. Качество определения плотности транспортного потока практически не изменяется при доле автомобилей с оборудованием больше 7%. При точности GPS приборов до 100 м относительная ошибка вычисления плотности потока уменьшается на 2% [1].

Работа поддержана грантом РФФИ № 14-07-00685.

- [1] *Петров Г. Е., Чехович Ю. В.* Идентификация имитационных моделей транспортных потоков с помощью разнородных источников прецедентной информации // Машинное обучение и анализ данных, 2015, №11. С. 1539–1544. jmla.org/papers/doc/2015/no11/Petrov2015Identification.pdf.

Identification of traffic flow simulation using dissimilar information sources

*Petrov Grigory*¹

greexon@gmail.com

Chekhovich Yury^{2*}

chehovich@forecsys.ru

¹Moscow, Lomonosov Moscow State University

²Moscow, Dorodnicyn Computing Centre of RAS

In the last few years, automobilization has grown rapidly. One of the most important problems is to collect information about traffic flows. Unfortunately, there do not exist absolutely full and reliable sources of this information. There are some approaches such as using traffic detectors, GPS technology, photocapture and video recording but they all have disadvantages. For example, traffic detectors determine traffic flow parameters in a limited part of road networks, GPS technology has low spatial accuracy and penetration rate, and photocapture and video recording depend on daylight and weather conditions.

In this paper, traffic flow simulation is used to understand bounds of data quality and data size needed to determine traffic flow parameters such as density and method of conducting the experiment. Single lane encircling highway is considered with installed traffic detectors and running vehicles. Many experiments are carried out with different values of GPS penetration rate and spatial accuracy. Traffic flow density is computed from vehicle speed by using Tanaka's, Greenshield's, and Greenberg's models.

According to experimental results, there is a clear bound of 7% of vehicles carrying the traffic monitoring equipment when the quality of determining traffic flow parameters such as density cannot be improved significantly with more penetration rate. In addition, GPS samples spatial accuracy of 100 m leads to 2 percent measurements relative error [1].

This research is funded by the Russian Foundation for Basic Research, grant 14-07-00685.

- [1] Petrov, G., and Chekhovich Y. 2015. Identification of traffic flow simulation using dissimilar information sources *J. Mach. Learn. Data Anal.* 1(11):1539–1554. jmla.org/papers/doc/2015/no11/Petrov2015Identification.pdf.

Содержание

Композиционные методы анализа данных	10
<i>Адуенко А. А., Стрижов В. В.</i>	
Анализ пространства параметров в задачах выбора мульти- моделей	10
<i>Иофина Г. В., Максимов Ю. В., Минаев А. В., Поляков Ю. С.</i>	
Вычислительно эффективные методы для обучения метри- к и усиленные оценки риска метрических классифика- торов	12
<i>Кириллов А. Н., Гавриков М. И., Лобачева Е. М., Осо- кин А. А., Ветров Д. П.</i>	
Многоклассовая модель формы со скрытыми переменными	14
<i>Крымова Е. А.</i>	
Агрегация упорядоченных оценок в цветном шуме	16
<i>Кузнецов М. П.</i>	
Комбинирование отношений порядка для восстановления предпочтения на наборе объектов	18
<i>Ланге М. М., Ганебных С. Н.</i>	
Многоклассовое распознавание в пространстве представ- лений с многоуровневым разрешением	20
<i>Ланге М. М., Ганебных С. Н., Ланге А. М.</i>	
Алгоритм приближенного поиска ближайшего цифрово- го массива на множестве пирамидальных представлений данных	22
<i>Маленичев А. А., Красоткина О. В., Моттль В. В.</i>	
Многоклассовое распознавание в больших массивах дан- ных	24
<i>Молчанов Д. А., Кондрашкин Д. А., Ветров Д. П.</i>	
Машина релевантных тегов	26
<i>Мотренко А. П.,</i>	
Оценка объема выборки в задачах классификации	28
<i>Неделько В. М.</i>	
К вопросу об эффективности бустинга в задаче класси- фикации	30

<i>Новиков А. В., Родоманов А. О., Осокин А. А., Ветров Д. П.</i> Тензорный поезд в марковском случайном поле	32
<i>Рязанов В. В., Виноградов А. П., Лаптин Ю. П.</i> Использование обобщенных прецедентов для сжатия больших выборок при обучении	34
<i>Сенько О. В., Морозов А. М., Кузнецова А. В., Клименко Л. Л.</i> Оценка эффекта множественного тестирования при поиске закономерностей в данных высокой размерности	36
<i>Торшин И. Ю., Рудаков К. В.</i> Об основах метрического анализа плохо формализованных задач распознавания и классификации	38
<i>Хачай М. Ю.</i> Бустинг приближенных алгоритмов в приложении к полиэдральной отделимости комитетного типа	40
<i>Швец М. Ю., Зухба А. В., Воронцов К. В.</i> Построение монотонного классификатора для задач медицинской диагностики	42
<i>Шибзухов З. М.</i> Агрегирующие корректные операции над алгоритмами	44
<i>Шибзухов З. М., Чередников Д. Ю.</i> О моделях нейронов агрегирующего типа	46
Линейные и метрические методы восстановления зависимостей	48
<i>Бериков В. Б.</i> Коллективный кластерный анализ с весами, основанный на коассоциативных матрицах	48
<i>Борисова И. А., Кутненко О. А.</i> Цензурирование ошибочно классифицированных объектов выборки	50
<i>Варфоломеева А. А., Бочкарев А. М.</i> Структурное обучение при порождении моделей	52
<i>Двоенко С. Д., Пшеничный Д. О.</i> О метрических свойствах медианы Кемени	54

<i>Квасов А. Ф., Красоткина О. В.</i> Применение байесовской регуляризации при обучении распознавания сигналов на основе метода опорных векторов	56
<i>Ларин А. О., Середин О. С., Моттль В. В.</i> Управление селективностью аномальных наблюдений в обучении одноклассовому распознаванию образов	58
<i>Рудой Г. И.</i> Исследование и выбор регрессионных моделей с учетом погрешности как зависимых, так и независимых переменных	60
<i>Сенько О. В., Докукин А. А.</i> Регрессионная модель, основанная на выпуклых комбинациях, максимально коррелирующих с откликом	62
<i>Середин О. С., Моттль В. В.</i> Метрическая гипотеза компактности и метод потенциалных функций в теории обучения	64
<i>Турков П. А., Красоткина О. В.</i> Отбор признаков в задаче классификации в потоках данных при смещении решающего правила	66
Дискретно-логические методы классификации	68
<i>Викентьев А. А., Фефелова В. В.</i> Новые расстояния и меры достоверности для формул логики Лукасевича в кластеризации логических высказываний базы знаний	68
<i>Дюкова Е. В., Журавлёв Ю. И., Прокофьев П. А.</i> Вопросы эффективности логических корректоров	70
<i>Дюкова Е. В., Никифоров А. Г.</i> Об эффективных параллельных асимптотически оптимальных алгоритмах дуализации	72
<i>Ильев В. П., Ильева С. Д.</i> Об одной задаче кластеризации на графе	74
<i>Ишкина Ш. Х.</i> Комбинаторные оценки переобучения одномерных пороговых классификаторов	76

<i>Кононов А. В., Симанчев Р. Ю., Уразова И. В.</i> Задача отделимости (separation problem) опорных неравенств для задачи аппроксимации графа	78
<i>Юдин В. Н., Карпов Л. Е., Карпов Ю. Л., Сметанин Ю. Г.</i> Дифференцирование классов при оценке не полностью описанного объекта	80
<i>Янковская А. Е., Ямшанов А. В., Кривдюк Н. М.</i> 2-симплекс призма — когнитивное средство принятия и обоснования решений в интеллектуальных динамических системах	82
Сложность вычислений и оптимизация	84
<i>Аникин А. С., Деуреченский П. Е., Гасников А. В., Горнов А. Ю., Камзолов Д. И., Максимов Ю. В.</i> Эффективные численные методы для дважды разреженных задач оптимизации большой размерности	84
<i>Бекларян Л. А., Хачатрян Н. К.</i> Динамическая модель организации грузоперевозок	86
<i>Гимади Э. Х.</i> О некоторых трудных задачах дискретной оптимизации, связанных с кластеризацией, покрытием и маршрутизацией	88
<i>Гимади Э. Х., Цидулко О. Ю.</i> Об асимптотически точных приближенных алгоритмах для некоторых трудных задач маршрутизации	90
<i>Гранин С. С., Максимов Ю. В.</i> Приближенные методы минимизации булевых функций в специальных классах	92
<i>Еремеев А. В.</i> Построение хабов для рынка электроэнергии	94
<i>Зухба А. В.</i> Отбор объектов и признаков для монотонных классификаторов	96

<i>Кельманов А. В.</i>	
Квадратичные евклидовы задачи 2-кластеризации: сложность и эффективные алгоритмы с оценками точности для их решения	98
<i>Кельманов А. В., Моткова А. В.</i>	
Точный псевдополиномиальный алгоритм для одной квадратичной евклидовой задачи сбалансированной 2-кластеризации	100
<i>Кельманов А. В., Пяткин А. В.</i>	
О сложности одной квадратичной евклидовой задачи сбалансированной 2-кластеризации	102
<i>Кельманов А. В., Хамидуллин С. А., Хандеев В. И.</i>	
Полностью полиномиальная приближенная схема для одной задачи 2-кластеризации последовательности	104
<i>Кельманов А. В., Хандеев В. И.</i>	
Полностью полиномиальная приближенная схема для одной квадратичной евклидовой задачи 2-кластеризации	106
<i>Лемтюжникова Д. В.</i>	
Модификации локального элиминационного алгоритма для эффективного решения больших разреженных задач дискретной оптимизации	108
<i>Максимов Ю. В., Решетова Д. Г.</i>	
Оптимистические оценки риска в задачах многоклассовой классификации	110
<i>Плясунов А. В., Кочетов Ю. А., Чвокич Д.</i>	
Существование равновесий в двухуровневых задачах ценообразования и размещения хабов	112
<i>Родманов А. О., Кропотов Д. А.,</i>	
Быстрый инкрементальный метод оптимизации больших сумм функций с суперлинейной скоростью сходимости	114
<i>Стенина М. М.</i>	
Согласование прогнозов в задачах прогнозирования иерархических временных рядов	116
<i>Теклина Л. Г., Котельников И. В.</i>	
Параметрический синтез робастных систем квазиинвариантного управления методами распознавания образов	118

Обработка изображений	120
<i>Визильтер Ю. В., Желтов С. Ю.</i>	
Математические морфологии на основе неусложняющих ϵ -фильтров	120
<i>Грачева И. А., Копылов А. В.</i>	
Быстрые алгоритмы обработки изображений на основе нестационарной гамма-нормальной модели	122
<i>Медведева Е. В., Трубин И. С., Устюжанина Е. А., Лалетин А. В.</i>	
Нелинейная многомерная фильтрация многокомпонентных изображений	124
<i>Новиков Е. А., Падалко М. А.</i>	
Использование Радон и Фурье преобразований растровых изображений для описания и отслеживания заданных объектов	126
<i>Петров Е. П., Харина Н. Л., Ржаникова Е. Д.</i>	
Комбинированная нелинейная фильтрация цифровых изображений большой разрядности	128
<i>Петров Е. П., Харина Н. Л., Сухих П. Н.</i>	
Метод быстрого сжатия изображений без потерь	130
<i>Рубис А. Ю., Лебедев М. А., Визильтер Ю. В.</i>	
Морфологическая фильтрация изображений на основе взаимного контрастирования	132
<i>Трекин А. Н., Матвеев И. А., Мурынин А. Б., Бочкарева В. Г.</i>	
Метод повышения разрешения космических изображений с использованием априорной информации в векторной форме для сохранения границ	134
<i>Федотов Н. Г., Сёмов А. А., Моисеев А. В.</i>	
Минимизация признакового пространства распознавания трехмерного изображения на основе стохастической геометрии и функционального анализа	136
<i>Харинов М. В.</i>	
Верификация модели квазиоптимальной сегментации на примере составного изображения	138

<i>Чочиа П. А.</i>	
Оценивание сложности изображений двумерными вариациями	140
Анализ изображений	142
<i>Кий К. И.</i>	
Сегментация и обнаружение контрастных объектов и их применение в навигации роботов	142
<i>Лебедев М. А., Рубис А. Ю., Горбачевич В. С., Визильтер Ю. В.</i>	
Выделение отличий на мозаичных изображениях на основе референтных фильтров	144
<i>Лихошерстный А. Ю., Жиляков Е. Г.</i>	
Обнаружение фрагментов изображений земной поверхности на основе субполосного анализа	146
<i>Мурашов Д. М., Березин А. В., Ганебных С. Н., Иванова Е. Ю.</i>	
Алгоритмы сегментации изображений сложной структуры	148
<i>Савченко А. В.</i>	
Посегментный анализ однородности в задаче распознавания изображений	150
Анализ видеопоследовательностей	152
<i>Вишняков Б. В., Горбачевич В. С., Сидякин С. В., Егоров А. И.</i>	
Быстрое межкадровое преобразование с использованием локальных бинарных паттернов	152
<i>Ипатов Ю. А., Парсаев Н. В., Тюкаев А. Ю.</i>	
Синтез алгоритма сегментации динамически изменяющихся сцен	154
<i>Князь В. В.</i>	
Распознавание эмоций, вызываемых музыкой, на основе анализа оптического потока изображений лица человека	156
<i>Князь В. А.</i>	
Методы распознавания и отслеживания меток в системе захвата движения	158

<i>Корнилов Ф. А.</i> Исследование влияния рассинхронизации входных изображений на качество работы алгоритмов поиска структурных различий	160
<i>Медведева Е. В., Карлушин К. А., Курбатова Е. Е.</i> Метод обнаружения движущихся объектов в видеопотоке на основе оценки их границ	162
<i>Нижибицкий Е. Ф.</i> Композиции признаков для видеотрекинга при помощи фильтра частиц	164
<i>Новиков Е. А., Ваколюк И. А.</i> Оптимизация метода компьютерной окулографии для исследований центральной нервной системы на основе пассивного анализа видеоизображения	166
<i>Цымбал Д. А., Чайкин А. В.</i> Вероятностный подход для отслеживания траекторий людей на видео с камер наблюдения	168
Распознавание изображений	170
<i>Горбачев В. С., Визильтер Ю. В., Воротников А. В.</i> Алгоритм последовательной жадной оптимизации целевой функции кодирования для создания биометрических шаблонов в задаче распознавания лиц	170
<i>Дробных К. А., Половинкин А. Н.</i> Использование алгоритмов глубокого обучения с учителем для оценки возраста человека по фотографии лица	172
Анализ формы изображений	174
<i>Ефимов Ю. С., Матвеев И. А.</i> Поиск внешней и внутренней границ радужной оболочки на изображении глаза методом парных градиентов	174
<i>Зенин Д. Г., Середин О. С., Кушнир О. А., Ларин А. О.</i> Комбинирование одноклассовой цветовой сегментации и скелетизации бинарных изображений для детектирования кисти руки	176

<i>Зюзин В. В., Поршнев С. В., Бобков А. О., Мухтаров А. А., Бобков В. В.</i>	
Анализ результатов оконтуривания левого желудочка сердца на эхографических изображениях у здоровых пациентов с помощью автоматического алгоритма	178
<i>Кушнир О. А., Середин О. С.</i>	
Определение зеркальной симметрии фигур на основе цепочек скелетных примитивов	180
<i>Местецкий Л. М., Тирас Х. П.</i>	
Нормализация изображений планарии методом жирных линий в биологических исследованиях	182
<i>Соломатин И. А., Матвеев И. А.</i>	
Определение видимой области радужки классификатором локальных текстурных признаков	184
<i>Флоринский И. В., Панкратов А. Н.</i>	
Цифровое моделирование рельефа с использованием ортогональных полиномов	186
<i>Чернышов В. Г., Местецкий Л. М.</i>	
Использование текстурного изображения внешней стороны пальцев в клиент-серверной системе биометрической идентификации	188
Анализ сигналов	190
<i>Анциперов В. Е.</i>	
Оценивание характера последствия случайных точечных процессов методами многомасштабного корреляционного анализа	190
<i>Власова К. В., Пахотин В. А., Клионский Д. М., Каплун Д. И.</i>	
Оценивание параметров радиоимпульсов методом максимального правдоподобия	192
<i>Жарких А. А.</i>	
Перспективы использования математических методов распознавания образов в технологиях встраивания цифровых водяных знаков в аудиосигнал	194

<i>Жарких А. А., Слащев А. А.</i> Анализ возможностей использования технологий data mining для изучения сложных сигналов и систем	196
<i>Рудаков К. В., Каширин Д. О., Стрижов В. В., Корчагин А. Д.</i> Разработка и исследование свойств системы прогнозирования объемов спроса на грузовые железнодорожные перевозки	198
<i>Чепрасов Д. Н., Маленичев А. А., Сулимова В. В., Красоткина О. В., Моттль В. В., Марков А. А.</i> Восстановление пропущенных данных на ультразвуковых дефектограммах рельсового пути на основе локально-глобального выравнивания	200
Приложения: биология и медицина	202
<i>Дорофеев Ю. А., Глазман Л. Ю., Дорофеев А. А., Мандель А. С., Потапов А. А., Чернявский А. Л., Шиманский В. Н., Шифрин М. А.</i> Разработка подсистемы маршрутизации пациентов в рамках интеллектуальной пациент-ориентированной системы управления крупной клиникой	202
<i>Махортыл С. А., Лыжко Е. В.</i> Локализация источников биомагнитной активности мозга	204
<i>Немирко А. П., Калинин А. Н., Манило Л. А.</i> Анализ стадий анестезии на основе оценки энтропии электроэнцефалограммы	206
<i>Панкратов А. Н., Руднев В. Р., Куликова Л. И., Дедус Ф. Ф., Тихонов Д. А., Ефимов А. В.</i> Конформационный анализ структурных мотивов типа α - α -уголок в вычислительном эксперименте молекулярной динамики	208

<i>Покровская И. В., Гучук В. В., Десова А. А., Дорофеев А. А., Дорофеев Ю. А.</i> Методы классификации и распознавания образов для анализа квазипериодических биосигналов в задачах ме- дицинской диагностики (на примере пульсового сигнала лучевой артерии)	210
<i>Рыкунов С. Д., Оплачко Е. С., Полянин А. Г., Устинин М. Н.</i> Технология создания атласа парциальных спектров отде- лов головного мозга человека	212
<i>Сушкова О. С., Габова А. В., Карабанов А. В., Кершнер И. А., Обухов К. Ю., Обухов Ю. В.</i> Метод частотно-временного анализа совместных измере- ний электроэнцефалограммы, электромиограммы и ме- ханического тремора при болезни Паркинсона	214
<i>Устинин М. Н., Рыкунов С. Д., Бойко А. И., Сычев В. В.</i> Реконструкция спонтанной активности головного мозга человека на основе анализа данных магнитной энцефа- лографии в пространстве «частота–паттерн»	216
<i>Целых В. Р., Воронцов К. В., Успенский В. М.</i> Сравнение информационного анализа электрокардиосиг- налов и анализа вариабельности сердечного ритма как методов диагностики заболеваний внутренних органов	218
Приложения: анализ текстов	220
<i>БазмUTOва И. В., Гусев В. Д., Мирошниченко Л. А., Тит- кова Т. Н.</i> Параллельные тексты в задаче дешифровки древнерус- ских знаменных песнопений	220
<i>Воронцов К. В., Фрей А. И., Апишев М. А., Дойков Н. В., Суворова М. А.</i> Регуляризация тематических моделей в библиотеке с от- крытым кодом BigARTM	222
<i>Кузьмин А. А., Адуенко А. А., Стрижов В. В.</i> Построение иерархических тематических моделей круп- ных конференций	224

<i>Михайлов Д. В., Козлов А. П., Емельянов Г. М.</i> Мера TF-IDF и формирование единиц представления знаний для открытых тестов	226
<i>Плавин А. В., Потапенко А. А., Воронцов К. В.</i> Энтропийный регуляризатор отбора тем в вероятност- ных тематических моделях	228
<i>Чиркова Н. А., Айсина Р. М., Воронцов К. В.</i> Иерархическая аддитивно регуляризованная тематиче- ская модель научной конференции	230
Прикладные системы	232
<i>Гуз И. С., Лексин В. А., Трофимов М. И., Фенстер А. М.</i> Эволюция подходов к модерации контента для сайта объ- явлений: от рекомендаций до автоматизации	232
<i>Дорофеюк А. А., Дорофеюк Ю. А., Покровская И. В., Чер- нявский А. Л.</i> Структурно-классификационные методы интеллектуаль- ного анализа данных для слабо формализованных систем управления	234
<i>Жуков А. В., Сидоров Д. Н., Томин Н. В., Курбацкий В. Г., Панасецкий Д. А., Спиряев В. А.</i> Модели на основе случайных лесов в задаче мониторинга состояний электроэнергетических систем	236
<i>Мамырбаев О., Калимолдаев М., Амиргалиев Е., Мусаба- ев Р., Алимхан К.</i> Метод разработки системы многомодального распозна- вания казахской речи	238
<i>Петров Г. Е., Чехович Ю. В.</i> Идентификация имитационных моделей транспортных потокос с помощью разнородных источников precedent- ной информации	240

Contents

Compositional Data Mining Methods	10
<i>Aduenko A., Strijov V.</i>	
Features space analysis for multimodel selection	11
<i>Iofina G., Maximov Yu., Minaev A., Polyakov Yu.</i>	
Computationally efficient metric learning with improved risk bounds	13
<i>Kirillov A., Gavrikov M., Lobacheva E., Osokin A., Vetrov D.</i>	
Deep part-based generative shape model with Latent variables	15
<i>Krymova E.</i>	
Aggregation of ordered smoothers in colored noise	17
<i>Kuznetsov M.</i>	
Combining order relations for estimating preferences over the set of objects	19
<i>Lange M., Ganebnykh S.</i>	
Multiclass pattern recognition in a space of the multiresolu- tion representations	21
<i>Lange M., Ganebnykh S., Lange A.</i>	
Algorithm of approximate search for the nearest digital array in a set of pyramidal data representations	23
<i>Malenichev A., Krasotkina O., Mottl V.</i>	
Multiclass learning in big data	25
<i>Molchanov D., Kondrashin D., Vetrov D.</i>	
Relevance tagging machine	27
<i>Motrenko A.</i>	
Sample size estimation in classification problems	29
<i>Nedelko N.</i>	
On performance of boosting in classification problem	31
<i>Novikov A., Rodomanov A., Osokin A., Vetrov D.</i>	
Putting Markov random fields on a tensor train	33
<i>Ryazanov V., Vinogradov A., Laptin Yu.</i>	
Using generalized precedents for big data sample compres- sion at learning	35

<i>Senko O., Morozov A., Kuznetsova A., Klimenko L.</i>	
Evaluating of multiple testing effect in task of regularities search in multidimensional data.	37
<i>Torshin I., Rudakov K.</i>	
Metric analysis of poorly formalized problems	39
<i>Khachay M.</i>	
Boosting of polynomial time approximation algorithms for committee polyhedral separability	41
<i>Shvets M., Zuhba A., Vorontsov K.</i>	
Learning isotonic classifier for medical diagnostics	43
<i>Shibzukhov Z.</i>	
Aggregation correct operations on algorithms	45
<i>Shibzukhov Z., Cherednikov D.</i>	
On the models of neurons of aggregation type	47
Linear and Metric Learning Methods	48
<i>Berikov V.</i>	
Ensemble clustering using weights based on coassociation matrices	49
<i>Borisova I., Kutnenko O.</i>	
Outliers detection in datasets with misclassified objects	51
<i>Varfolomeeva A., Bochkarev A.</i>	
Structured learning for model generation	53
<i>Dvoenko S., Pshenichny D.</i>	
On metric characteristics of the Kemeny's median	55
<i>Kvasov A., Krasotkina O.</i>	
Application of Bayesian regularization in signal recognition based on Support Vector Machine	57
<i>Larin A., Seregin O., Mottl V.</i>	
Outlier selectivity control in one-class pattern recognition learning	59
<i>Rudoy G.</i>	
Model selection and analysis with respect to measurement errors in both free and dependent variables	61

<i>Senko O., Dokukin A.</i>	
Regression model based on convex combinations best correlated with response	63
<i>Seredin O., Mottl V.</i>	
The metric-based compactness hypothesis and method of potential functions for machine learning	65
<i>Turkov P., Krasotkina O.</i>	
Feature selection in the data stream classification under concept drift	67
Rule-Based Classification	68
<i>Vikent'ev A., Fefelova V.</i>	
New distances and invalidation measures for formulas of Lukasiewicz logic in clusterization of logical propositions of a database	69
<i>Djukova E., Zhuravlev Yu., Prokofjev P.</i>	
The effectiveness of logical correctors	71
<i>Djukova E., Nikiforov A.</i>	
On efficient parallel asymptotically optimal dualization algorithms	73
<i>Il'ev V., Il'eva S.</i>	
On a graph clustering problem	75
<i>Ishkina Sh.</i>	
Combinatorial bounds of overfitting for threshold decision rules	77
<i>Kononov A., Simanchev R., Urazova I.</i>	
On separation problem of same support inequalities class for M -graphs polytope	79
<i>Yudin V., Karpov L., Karpov Yu., Smetanin Yu.</i>	
Differential set of classes for estimation of not fully described objects	81
<i>Yankovskaya A., Yamshanov A., Krivdyuk N.</i>	
2-simplex prism — a cognitive tool for decision-making and its justifications in intelligent dynamic systems	83

Computational Complexity and Optimization	84
<i>Anikin A., Dvurechensky P., Gasnikov A., Gornov A., Kamzolov D., Maximov Yu.</i>	
Efficient numerical optimization methods for huge-scale sparse problems	85
<i>Beklaryan L., Khachatryan N.</i>	
Dynamic model of the organization of cargo transportation .	87
<i>Gimadi E.</i>	
On some hard discrete optimization problems associated with clustering, covering, and rooting	89
<i>Gimadi E., Tsidulko O.</i>	
On polynomial asymptotically optimal algorithms for some hard routing problems	91
<i>Granin S., Maximov Yu.</i>	
Approximate minimization of disjunctive normal forms of Boolean functions from special classes	93
<i>Eremeev A.</i>	
Trading Hubs Construction for Electricity Markets	95
<i>Zukhba A.</i>	
Feature and prototype selection for isotonic classifier	97
<i>Kel'manov A.</i>	
Quadratic Euclidian 2-clustering problems: Complexity and efficient algorithms with performance guarantees for their solutions	99
<i>Kel'manov A., Motkova A.</i>	
An exact pseudopolynomial algorithm for an Euclidean balanced variance-based 2-clustering problem	101
<i>Kel'manov A., Pyatkin A.</i>	
On the complexity of Quadratic Euclidean Balanced 2-Clustering problem	103
<i>Kel'manov A., Khamidullin S., Khandeev V.</i>	
Fully polynomial-time approximation scheme for a sequence 2-clustering problem	105
<i>Kel'manov A., Khandeev V.</i>	
Fully polynomial-time approximation scheme for a quadratic Euclidean 2-clustering problem	107

<i>Lemtyuzhnikova D.</i>	
Modification of local elimination algorithm for sparse large-scale problems of discrete optimization	109
<i>Maximov Yu., Reshetova D.</i>	
Optimistic risk bounds in multiclass classification problems	111
<i>Plyasunov A., Kochetov Yu., Čvokić D.</i>	
Existence of equilibria in the bilevel hub location and pricing problem	113
<i>Rodomanov A., Kropotov D.</i>	
Fast incremental optimization of large sums with superlinear convergence rate	115
<i>Stenina M.</i>	
Forecasts reconciliation for hierarchical time series forecasting problem	117
<i>Teklina L., Kotel'nikov I.</i>	
Parametric synthesis of the robust quasi-invariant control systems using methods of pattern recognition	119
Image Processing	120
<i>Vizilter Yu., Zheltov S.</i>	
Mathematical morphologies based on nondecreasing ϵ -filters	121
<i>Gracheva I., Kopylov A.</i>	
Fast image processing algorithms on the basis of nonstationary gamma-normal model	123
<i>Medvedeva E., Trubin I., Ustyuzhanina E., Laletin A.</i>	
Multidimensional nonlinear filtration of multicomponent images	125
<i>Novikov E., Padalko M.</i>	
Using Radon and Fourier transformations of raster images for description and tracking of specified objects	127
<i>Petrov E., Kharina N., Rzhanikova E.</i>	
Combined nonlinear filtration of digital high bitness images	129
<i>Petrov E., Kharina N., Sukhikh P.</i>	
Fast lossless image compression method	131

<i>Rubis A., Lebedev M., Vizilter Yu.</i> Morphological image filtration based on mutual contrast . . .	133
<i>Trekin A., Matveev I., Murynin A., Bochkareva V.</i> A method for upsampling of remote sensing images using vector data for preserving edges	135
<i>Fedotov N., Syemov A., Moiseev A.</i> Feature space minimization of three-dimensional image recognition based on stochastic geometry and functional analysis	137
<i>Kharinov M.</i> Verification of quasi-optimal segmentation model using com- posite images	139
<i>Chochia P.</i> Image complexity assessment by two-dimensional variations	141
Image Analysis	142
<i>Kiy K.</i> Segmentation and detection of contrast objects and their application in robot navigation	143
<i>Lebedev M., Rubis A., Gorbatshevich V., Vizilter Yu.</i> Detecting differences on mosaic images based on reference fil- ters	145
<i>Likhosherstny A., Zhilyakov Y.</i> The discovery of fragments of images of the Earth based on surface subband analysis	147
<i>Murashov D., Berezin A., Ganebnykh S., Ivanova Y.</i> Techniques for segmenting images with a complex structure	149
<i>Savchenko A.</i> Segment homogeneity testing in the problem of image recog- nition	151
Video Analysis	152
<i>Vishnyakov B., Gorbatshevich V., Sidyakin S., Egorov A.</i> Fast interframe transformation with local binary patterns . .	153

<i>Ipatov Y., Parsaev N., Tyukaev A.</i>	
Synthesis segmentation algorithm for dynamically changing scenes	155
<i>Kniaz V.</i>	
Music-elicited emotion identification using optical flow analysis of human face	157
<i>Knyaz V.</i>	
Methods for recognizing and tracing targets for motion capture system	159
<i>Kornilov F.</i>	
Investigation of the impact of misregistration of input images on the accuracy of change detection	161
<i>Medvedeva E., Karlushin K., Kurbatova E.</i>	
Method of detection of moving objects in video stream on the basis of object boundaries estimation	163
<i>Nizhibitsky E.</i>	
Feature composition in video tracking using particle filters .	165
<i>Novikov E., Vakoliuk I.</i>	
Optimization method of computer oculography for research of the central nervous system based on passive video analysis	167
<i>Tsybal D., Chaykin A.</i>	
Probabilistic approach to people tracking in surveillance video	169
Image Recognition	170
<i>Gorbatsevich V., Vizilter Yu., Vorotnikov A.</i>	
Sequential greedy optimization of coding function for binary templates construction in face recognition	171
<i>Drobnyh K., Polovinkin A.</i>	
Using Unsupervised Deep Learning for Human Age Estimation Problem	173
Images Shape Analysis	174
<i>Efimov Y., Matveev I.</i>	
Locating external and internal borders of the iris in eye image by paired gradients method	175

<i>Zenin D., Seredin O., Kushnir O., Larin A.</i>	
Combining of one-class color segmentation and binary images skeletonization for hand detecting	177
<i>Zyuzin V., Porshnev S., Bobkova A., Mukhtarov A., Bobkov V.</i>	
The analysis of results of the left ventricle contouring using automatic algorithm on ultrasound images	179
<i>Kushnir O., Seredin O.</i>	
Reflectional symmetry of shape based on skeleton chains of primitives	181
<i>Mestetskiy L., Tiras Kh.</i>	
Normalization of planarian images by fat curves in biological studies	183
<i>Solomatin I., Matveev I.</i>	
Detecting visible areas of iris by qualifier of local textural features	185
<i>Florinsky I., Pankratov A.</i>	
Digital terrain modeling with orthogonal polynomials	187
<i>Chernyshov V., Mestetskiy L.</i>	
Using finger knuckle prints in client-server biometric system	189
Signal Analysis	190
<i>Antsiperov V.</i>	
Estimation of renewal point process aftereffect characteristics on the bases of multiscale correlation analysis methods.	191
<i>Vlasova K., Pachotin V., Klionskiy D., Kaplun D.</i>	
Estimation of radio impulse parameters using the maximum likelihood method	193
<i>Zharkikh A.</i>	
Perspectives of uses of mathematical methods of image identification in technologies of embedding of digital water marks in an audio signal	195
<i>Zharkikh A., Slashchev A.</i>	
Analysis of possibilities in using data mining technologies to study complex signals and systems	197

<i>Rudakov K., Kashirin D., Strijov V., Korchagin A.</i> Design and analysis of a forecasting system for rail freight transportation demand	199
<i>Cheprasov D., Malenichev A., Sulimova V., Krasotkina O., Mottl V., Markov A.</i> Recovering missing data on ultrasonic rail defectograms via semiglobal warping	201
Applications: Medicine and Biology	202
<i>Dorofeyuk Yu., Glazman L., Dorofeyuk A., Mandel A., Potapov A., Chernyavskiy A., Shimansky V., Shifrin M.</i> Development of subsystem routing patients through intelli- gent patient-oriented system for the management of large clinic	203
<i>Makhortykh S., Lyzhko E.</i> Sources localization for brain biomagnetic activity	205
<i>Nemirko A., Kalinichenko A., Manilo L.</i> Analysis of anesthesia stages based on the electroen- cephalography entropy estimation	207
<i>Pankratov A., Rudnev V., Kulikova L., Dedus F., Tikhonov D., Efimov A.</i> Conformational analysis of structural motifs of α - α -corner in the computational experiment of molecular dynamics	209
<i>Pokrovskaya I., Guchuk V., Desova A., Dorofeyuk A., Doro- feyuk Yu.</i> Methods of classification and pattern recognition for the analysis of quasi-periodic biosignals in medical diagnostics (on an example of radial artery pulse signal)	211
<i>Rykunov S., Oplachko E., Polyanin A., Ustinin M.</i> Technology to build the atlas of the partial spectra of the hu- man brain	213
<i>Sushkova O., Gabova A., Karabanov A., Kershner I., Obukhov K., Obukhov Yu.</i> A method of time-frequency analysis of joint electroen- cephalogram, electromyogram, and mechanical tremor sig- nals in Parkinson's disease	215

<i>Ustinin M., Rykunov S., Boyko A., Sychev V.</i> Reconstruction of human brain spontaneous activity based on frequency–pattern analysis of magnetoencephalography data	217
<i>Tselykh V., Vorontsov K., Uspenskiy V.</i> The comparison of the information analysis of ECG-signals and the heart rate variability analysis as techniques for the diagnostics of internal diseases	219
Applications: Text Mining	220
<i>Bakhmutova I., Gusev V., Miroshnichenko L., Titkova T.</i> Parallel texts in the problem of deciphering of ancient Rus- sian Chant	221
<i>Vorontsov K., Frei O., Apishev M., Doikov N., Suvorova M.</i> Regularization of topic models in BigARTM open source library	223
<i>Kuzmin A., Aduenko A., Strijov V.</i> Hierarchical thematic modeling of major conference proceed- ings	225
<i>Mikhaylov D., Kozlov A., Emelyanov G.</i> TF-IDF metrics and formation of units for knowledge rep- resentation in open tests	227
<i>Plavin A., Potapenko A., Vorontsov K.</i> Topic selection in probabilistic topics models with entropy regularization	229
<i>Chirkova N., Aysina R., Vorontsov K.</i> Hierarchical additively regularized topic model of a scientific conference	231
Application Systems	232
<i>Guz I., Leksin V., Trofimov M., Fenster A.</i> Evolution of content moderation approaches for online clas- sifieds: From action recommendations to automation	233
<i>Dorofeyuk A., Dorofeyuk Yu., Pokrovskaya I., Chernyavskiy A.</i> Structural classification data mining techniques for weakly formalized control systems	235

<i>Zhukov A., Sidorov D., Tomin N., Kurbatsky V., Panaset- sky D., Spiryaev V.</i>	
Random forests based models for emergency state monitor- ing in power systems	237
<i>Mamyrbayev O., Kalimoldaev M., Amirgaliyev E., Mus- abayev R., Alimhan K.</i>	
Development method of multimodal recognition of the Kazakh speech	239
<i>Petrov G., Chekhovich Yu.</i>	
Identification of traffic flow simulation using dissimilar in- formation sources	241

Авторский указатель**А**

Адуенко А. А. 10, 224
Айсина Р. М. 230
Алимхан К. 238
Амиргалиев Е. 238
Аникин А. С. 84
Анциперов В. Е. 190
Апишев М. А. 222

Б

Бахмутова И. В. 220
Бекларян Л. А. 86
Березин А. В. 148
Бериков В. Б. 48
Бобков А. О. 178
Бобков В. В. 178
Бойко А. И. 216
Борисова И. А. 50
Бочкарев А. М. 52
Бочкарева В. Г. 134

В

Ваколюк И. А. 166
Варфоломеева А. А. 52
Ветров Д. П. 14, 26, 32
Визильтер Ю. В. ... 120, 132,
144, 170
Викентьев А. А. 68
Виноградов А. П. 34
Вишняков Б. В. 152
Власова К. В. 192
Воронцов К. В. 42, 218, 222,
228, 230
Воротников А. В. 170

Г

Габова А. В. 214
Гавриков М. И. 14
Ганебных С. Н. ... 20, 22, 148
Гасников А. В. 84
Гимади Э. Х. 88, 90
Глазман Л. Ю. 202
Горбацевич В. С. ... 144, 152,
170
Горнов А. Ю. 84
Гранин С. С. 92
Грачева И. А. 122
Гуз И. С. 232
Гусев В. Д. 220
Гучук В. В. 210

Д

Двоенко С. Д. 54
Двуреченский П. Е. 84
Дедус Ф. Ф. 208
Десова А. А. 210
Дойков Н. В. 222
Докукин А. А. 62
Дорофеюк А. А. ... 202, 210,
234
Дорофеюк Ю. А. ... 202, 210,
234
Дробных К. А. 172
Дюкова Е. В. 70, 72

Е

Егоров А. И. 152
Емельянов Г. М. 226
Еремеев А. В. 94
Ефимов А. В. 208
Ефимов Ю. С. 174

Ж

Жарких А. А. 194, 196
 Желтов С. Ю. 120
 Жилияков Е. Г. 146
 Жуков А. В. 236
 Журавлёв Ю. И. 70

З

Зенин Д. Г. 176
 Зухба А. В. 42, 96
 Зюзин В. В. 178

И

Иванова Е. Ю. 148
 Ильев В. П. 74
 Ильева С. Д. 74
 Иофина Г. В. 12
 Ипатов Ю. А. 154
 Ишкина Ш. Х. 76

К

Калимолдаев М. 238
 Калиниченко А. Н. 206
 Камзолов Д. И. 84
 Каплун Д. И. 192
 Карабанов А. В. 214
 Карлушин К. А. 162
 Карпов Л. Е. 80
 Карпов Ю. Л. 80
 Каширин Д. О. 198
 Квасов А. Ф. 56
 Кельманов А. В. 98, 100,
 102, 104, 106
 Кершнер И. А. 214
 Кий К. И. 142
 Кириллов А. Н. 14
 Клименко Л. Л. 36
 Клионский Д. М. 192

Князь В. А. 158
 Князь В. В. 156
 Козлов А. П. 226
 Кондрашкин Д. А. 26
 Кононов А. В. 78
 Копылов А. В. 122
 Корнилов Ф. А. 160
 Корчагин А. Д. 198
 Котельников И. В. 118
 Кочетов Ю. А. 112
 Красоткина О. В. 24, 56, 66,
 200
 Кривдюк Н. М. 82
 Кропотов Д. А. 114
 Крымова Е. А. 16
 Кузнецов М. П. 18
 Кузнецова А. В. 36
 Кузьмин А. А. 224
 Куликова Л. И. 208
 Курбатова Е. Е. 162
 Курбацкий В. Г. 236
 Кутненко О. А. 50
 Кушнир О. А. 176, 180

Л

Лалетин А. В. 124
 Ланге А. М. 22
 Ланге М. М. 20, 22
 Лаптин Ю. П. 34
 Ларин А. О. 58, 176
 Лебедев М. А. 132, 144
 Лексин В. А. 232
 Лемтюжникова Д. В. ... 108
 Лихошерстный А. Ю. ... 146
 Лобачева Е. М. 14
 Лыжко Е. В. 204

М

- Максимов Ю. В. . . 12, 84, 92,
110
Маленичев А. А. 24, 200
Мамырбаев О. 238
Мандель А. С. 202
Манило Л. А. 206
Марков А. А. 200
Матвеев И. А. . . 134, 174, 184
Махортых С. А. 204
Медведева Е. В. . . . 124, 162
Местецкий Л. М. . . . 182, 188
Минаев А. В. 12
Мирошниченко Л. А. . . . 220
Михайлов Д. В. 226
Моисеев А. В. 136
Молчанов Д. А. 26
Морозов А. М. 36
Моткова А. В. 100
Мотренко А. П. 28
Моттль В. В. . . 24, 58, 64, 200
Мурашов Д. М. 148
Мурынин А. Б. 134
Мусабаев Р. 238
Мухтаров А. А. 178

Н

- Неделько В. М. 30
Немирко А. П. 206
Нижибицкий Е. Ф. 164
Никифоров А. Г. 72
Новиков А. В. 32
Новиков Е. А. 126, 166

О

- Обухов К. Ю. 214
Обухов Ю. В. 214
Оплачко Е. С. 212
Осокин А. А. 14, 32

П

- Падалко М. А. 126
Панасецкий Д. А. 236
Панкратов А. Н. . . . 186, 208
Парсаев Н. В. 154
Пахотин В. А. 192
Петров Г. Е. 240
Петров Е. П. 128, 130
Плавин А. В. 228
Плясунов А. В. 112
Покровская И. В. . . . 210, 234
Половинкин А. Н. 172
Поляков Ю. С. 12
Полянин А. Г. 212
Поршнев С. В. 178
Потапенко А. А. 228
Потапов А. А. 202
Прокофьев П. А. 70
Пшеничный Д. О. 54
Пяткин А. В. 102

Р

- Решетова Д. Г. 110
Ржаникова Е. Д. 128
Родоманов А. О. 32, 114
Рубис А. Ю. 132, 144
Рудаков К. В. 38, 198
Руднев В. Р. 208
Рудой Г. И. 60
Рыкунов С. Д. 212, 216
Рязанов В. В. 34

С

- Савченко А. В. 150
Сенько О. В. 36, 62
Середин О. С. . . . 58, 64, 176,
180
Сёмов А. А. 136

Сидоров Д. Н. 236
 Сидякин С. В. 152
 Симанчев Р. Ю. 78
 Слащев А. А. 196
 Сметанин Ю. Г. 80
 Соломатин И. А. 184
 Спириев В. А. 236
 Стенина М. М. 116
 Стрижов В. В. .. 10, 198, 224
 Суворова М. А. 222
 Сулимова В. В. 200
 Сухих П. Н. 130
 Сушкова О. С. 214
 Сычев В. В. 216

Т

Теклина Л. Г. 118
 Тирас Х. П. 182
 Титкова Т. Н. 220
 Тихонов Д. А. 208
 Томин Н. В. 236
 Торшин И. Ю. 38
 Трекин А. Н. 134
 Трофимов М. И. 232
 Трубин И. С. 124
 Турков П. А. 66
 Тюкаев А. Ю. 154

У

Уразова И. В. 78
 Успенский В. М. 218
 Устинин М. Н. 212, 216
 Устюжанина Е. А. 124

Ф

Федотов Н. Г. 136
 Фенстер А. М. 232
 Фефелова В. В. 68

Флоринский И. В. 186
 Фрей А. И. 222

Х

Хамидуллин С. А. 104
 Хандеев В. И. 104, 106
 Харина Н. Л. 128, 130
 Харинов М. В. 138
 Хачай М. Ю. 40
 Хачатрян Н. К. 86

Ц

Целых В. Р. 218
 Цидулко О. Ю. 90
 Цымбал Д. А. 168

Ч

Чайкин А. В. 168
 Чвокич Д. 112
 Чепрасов Д. Н. 200
 Чередников Д. Ю. 46
 Чернышов В. Г. 188
 Чернявский А. Л. .. 202, 234
 Чехович Ю. В. 240
 Чиркова Н. А. 230
 Чочиа П. А. 140

Ш

Швец М. Ю. 42
 Шибзухов З. М. 44, 46
 Шиманский В. Н. 202
 Шифрин М. А. 202

Ю

Юдин В. Н. 80

Я

Ямшанов А. В. 82
 Янковская А. Е. 82

Author index**A**

Aduenko A. 11, 225
Alimhan K. 239
Amirgaliyev E. 239
Anikin A. 85
Antsiperov V. 191
Apishev M. 223
Aysina R. 231

B

Bakhmutova I. 221
Beklaryan L. 87
Berezin A. 149
Berikov V. 49
Bobkov V. 179
Bobkova A. 179
Bochkarev A. 53
Bochkareva V. 135
Borisova I. 51
Boyko A. 217

C

Chaykin A. 169
Chekhovich Yu. 241
Cheprasov D. 201
Cherednikov D. 47
Chernyavskiy A. 203, 235
Chernyshov V. 189
Chirkova N. 231
Chochia P. 141
Čvokić D. 113

D

Dedus F. 209
Desova A. 211
Djukova E. 71, 73

Doikov N. 223
Dokukin A. 63
Dorofeyuk A. 203, 211, 235
Dorofeyuk Yu. 203, 211, 235
Drobnyh K. 173
Dvoenko S. 55
Dvurechensky P. 85

E

Efimov A. 209
Efimov Y. 175
Egorov A. 153
Emelyanov G. 227
Eremeev A. 95

F

Fedotov N. 137
Fefelova V. 69
Fenster A. 233
Florinsky I. 187
Frei O. 223

G

Gabova A. 215
Ganebnykh S. 21, 23, 149
Gasnikov A. 85
Gavrikov M. 15
Gimadi E. 89, 91
Glazman L. 203
Gorbatsevich V. 145, 153, 171
Gornov A. 85
Gracheva I. 123
Granin S. 93
Guchuk V. 211
Gusev V. 221
Guz I. 233

I

Il'ev V.75
 Il'eva S.75
 Iofina G.13
 Ipatov Y.155
 Ishkina Sh.77
 Ivanova Y.149

K

Kalimoldaev M.239
 Kalinichenko A.207
 Kamzolov D.85
 Kaplun D.193
 Karabanov A.215
 Karlushin K.163
 Karpov L.81
 Karpov Yu.81
 Kashirin D.199
 Kel'manov A. ..99, 101, 103,
 105, 107
 Kershner I.215
 Khachatryan N.87
 Khachay M.41
 Khamidullin S.105
 Khandeev V.105, 107
 Kharina N.129, 131
 Kharinov M.139
 Kirillov A.15
 Kiy K.143
 Klimenko L.37
 Klionskiy D.193
 Kniaz V.157
 Knyaz V.159
 Kochetov Yu.113
 Kondrashin D.27
 Kononov A.79
 Kopylov A.123
 Korchagin A.199

Kornilov F.161
 Kotel'nikov I.119
 Kozlov A.227
 Krasotkina O. 25, 57, 67, 201
 Krivdyuk N.83
 Kropotov D.115
 Krymova E.17
 Kulikova L.209
 Kurbatova E.163
 Kurbatsky V.237
 Kushnir O.177, 181
 Kutnenko O.51
 Kuzmin A.225
 Kuznetsov M.19
 Kuznetsova A.37
 Kvasov A.57

L

Laletin A.125
 Lange A.23
 Lange M.21, 23
 Laptin Yu.35
 Larin A.59, 177
 Lebedev M.133, 145
 Leksin V.233
 Lemtyuzhnikova D.109
 Likhosherstny A.147
 Lobacheva E.15
 Lyzhko E.205

M

Makhortykh S.205
 Malenichev A.25, 201
 Mamyrbayev O.239
 Mandel A.203
 Manilo L.207
 Markov A.201
 Matveev I.135, 175, 185
 Maximov Yu. .13, 85, 93, 111

Medvedeva E. 125, 163
Mestetskiy L. 183, 189
Mikhaylov D. 227
Minaev A. 13
Miroshnichenko L. 221
Moiseev A. 137
Molchanov D. 27
Morozov A. 37
Motkova A. 101
Motrenko A. 29
Mottl V. 25, 59, 65, 201
Mukhtarov A. 179
Murashov D. 149
Murynin A. 135
Musabayev R. 239

N

Nedelko N. 31
Nemirko A. 207
Nikiforov A. 73
Nizhibitsky E. 165
Novikov A. 33
Novikov E. 127, 167

O

Obukhov K. 215
Obukhov Yu. 215
Oplachko E. 213
Osokin A. 15, 33

P

Pachotin V. 193
Padalko M. 127
Panasetzky D. 237
Pankratov A. 187, 209
Parsaev N. 155
Petrov E. 129, 131
Petrov G. 241
Plavin A. 229

Plyasunov A. 113
Pokrovskaya I. 211, 235
Polovinkin A. 173
Polyakov Yu. 13
Polyanin A. 213
Porshnev S. 179
Potapenko A. 229
Potapov A. 203
Prokofjev P. 71
Pshenichny D. 55
Pyatkin A. 103

R

Reshetova D. 111
Rodomanov A. 33, 115
Rubis A. 133, 145
Rudakov K. 39, 199
Rudnev V. 209
Rudoy G. 61
Ryazanov V. 35
Rykunov S. 213, 217
Rzhanikova E. 129

S

Savchenko A. 151
Senko O. 37, 63
Seredin O. 59, 65, 177, 181
Shibzukhov Z. 45, 47
Shifrin M. 203
Shimansky V. 203
Shvets M. 43
Sidorov D. 237
Sidyakin S. 153
Simanchev R. 79
Slashchev A. 197
Smetanin Yu. 81
Solomatin I. 185
Spiryaev V. 237
Stenina M. 117

Strijov V. 11, 199, 225
 Sukhikh P. 131
 Sulimova V. 201
 Sushkova O. 215
 Suvorova M. 223
 Sychev V. 217
 Syemov A. 137

T

Teklina L. 119
 Tikhonov D. 209
 Tiras Kh. 183
 Titkova T. 221
 Tomin N. 237
 Torshin I. 39
 Trekin A. 135
 Trofimov M. 233
 Trubin I. 125
 Tselykh V. 219
 Tsidulko O. 91
 Tsymbal D. 169
 Turkov P. 67
 Tyukaev A. 155

U

Urazova I. 79
 Uspenskiy V. 219
 Ustinin M. 213, 217
 Ustyuzhanina E. 125

V

Vakoliuk I. 167
 Varfolomeeva A. 53
 Vetrov D. 15, 27, 33
 Vikent'ev A. 69
 Vinogradov A. 35
 Vishnyakov B. 153
 Vizilter Yu. ... 121, 133, 145,
 171
 Vlasova K. 193
 Vorontsov K. ... 43, 219, 223,
 229, 231
 Vorotnikov A. 171

Y

Yamshanov A. 83
 Yankovskaya A. 83
 Yudin V. 81

Z

Zenin D. 177
 Zharkikh A. 195, 197
 Zheltov S. 121
 Zhilyakov Y. 147
 Zhukov A. 237
 Zhuravlev Yu. 71
 Zukhba A. 43, 97
 Zyuzin V. 179

Научное издание

МАТЕМАТИЧЕСКИЕ МЕТОДЫ
РАСПОЗНАВАНИЯ ОБРАЗОВ

Тезисы докладов
17-й Всероссийской конференции
с международным участием

Напечатано с готового оригинал-макета

Сдано в набор 22.07.15. Подписано в печать 07.09.15.
Формат 60×90/16. Бумага офсетная. Печать цифровая.
Усл.-печ. л. 17,13. Уч.-изд. л. 13,0. Тираж 150 экз.

Заказ № 6909.

Издательство «ТОРУС ПРЕСС»
121614, г. Москва, ул. Крылатская 29-1-43
e-mail: torus@torus-press.ru
<http://www.torus-press.ru>



Отпечатано в цифровой типографии «Буки Веди»
на оборудовании Konica Minolta
119049, г. Москва, Ленинский проспект, д. 4, стр. 1А
Тел.: (495) 926-63-96, www.bukivedi.com, info@bukivedi.com

ISBN 978-5-94588-188-4



