

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

ПРИ ПОДДЕРЖКЕ

РОССИЙСКОГО ФОНДА ФУНДАМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ,
КОМПАНИЙ ФОРЕКСИС И ЦСПИР

Математические методы распознавания образов

ММРО-2017

Тезисы докладов
18-й Всероссийской конференции
с международным участием

ТОРУС
ПРЕСС  МОСКВА
2017

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 18-й Всероссийской конференции с международным участием, г. Таганрог, 2017 г. — М.: ТОРУС ПРЕСС, 2017. — 228 с.
ISBN 978-5-94588-225-6

В сборнике представлены тезисы докладов 18-й Всероссийской конференции «Математические методы распознавания образов», проводимой Вычислительным центром им. А. А. Дородницына ФИЦ ИУ РАН, Московским физико-техническим институтом и Южным федеральным университетом.

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-2017 выражает особую благодарность РФФИ и компаниям Форексис и ЦСПиР, оказавшим неоценимую помощь при подготовке и проведении конференции. Мероприятие проводится при финансовой поддержке РФФИ, проект № 17-07-20550.

ББК 22.1:32.973.26-018.2

ISBN 978-5-94588-225-6

© Авторы докладов, 2017

© ФИЦ ИУ РАН, 2017

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 18th All-Russian Conference with International Participation, Taganrog, 2017. — Moscow: TORUS PRESS, 2017. — 228 p. ISBN 978-5-94588-225-6

The volume contains the abstracts of the 18th 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 Southern Federal University.

The conference has being 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-2017 is grateful to the Russian Foundation for Basic Research, Forecsys Co. and CFRS Co. for providing assistance in the conference preparation and execution. The publication of the volume is funded by RFBR, grant 17-07-20550.

BBK 22.1:32.973.26-018.2

ISBN 978-5-94588-225-6

© Authors of the abstracts, 2017

© FRC CSC RAS, 2017

Оргкомитет

Председатель: Журавлев Юрий Иванович, *академик РАН*

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

Агеева Арина Андреевна
Борисова Татьяна Игоревна
Бунакова Влада Руслановна
Веселов Геннадий Евгеньевич, *д.т.н.*
Громов Андрей Николаевич
Инякин Андрей Сергеевич, *к.ф.-м.н.*
Карапетянц Алексей Николаевич, *д.ф.-м.н.*
Мотренко Анастасия Петровна
Помазкова Евгения Владимировна
Рейер Иван Александрович, *к.т.н.*
Рогозов Юрий Иванович, *д.т.н.*

Программный комитет

Сопредседатели: Рудаков Константин Владимирович, *академик РАН*

Зорин Денис Николаевич, *профессор*

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.*
Гупал Анатолий Михайлович, *д.ф.-м.н.*
Двоенко Сергей Данилович, *д.ф.-м.н.*
Дорофеюк Александр Александрович, *д.т.н.*
Каркищенко Александр Николаевич, *д.ф.-м.н.*
Кельманов Александр Васильевич, *д.ф.-м.н.*
Краснопрошин Виктор Владимирович, *д.т.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.*
Рязанов Владимир Васильевич, *д.ф.-м.н.*
Сойфер Виктор Александрович, *чл.-корр. РАН*
Тузиков Александр Васильевич, *д.ф.-м.н.*
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.*

Organizing Committee

Chair: Yury Zhuravlev, *acad. of RAS*

Secretary: Yury Chekhovich, *C.Sc.*

Arina Ageeva
Tatyana Borisova
Vlada Bunakova
Andrey Gromov
Andrey Inyakin, *C.Sc.*
Alexey Karapetiants, *D.Sc.*
Anastasia Motrenko
Evgeniya Pomazkova
Ivan Reyer, *C.Sc.*
Yury Rogozov, *D.Sc.*
Gennady Veselov, *D.Sc.*

Program Committee

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

Secretary: Vadim Strijov, *D.Sc.*

Alexey Chulichkov, *D.Sc.*
Alexander Dorofeyuk, *D.Sc.*
Sergey Dvoenko, *D.Sc.*
Edward Gimadi, *D.Sc.*
Alexander Gornov, *D.Sc.*
Olga Gromova, *D.Sc.*
Anatoly Gupal, *D.Sc.*
Alexander Karkishchenko, *D.Sc.*
Alexander Kel'manov, *D.Sc.*
Michael Khachay, *D.Sc.*
Victor Krasnoproshin, *D.Sc.*
Leonid Mestetskiy, *D.Sc.*
Vadim Mottl, *D.Sc.*
Gennady Osipov, *D.Sc.*
Yury Pytiev, *D.Sc.*
Vladimir Ryazanov, *D.Sc.*
Alexander Shanenin, *D.Sc.*
Viktor Soyfer, *corr. member of RAS*
Alexander Tuzikov, *D.Sc.*
Mikhail Ustinin, *D.Sc.*
Konstantin Vorontsov, *D.Sc.*

Рецензенты

Адуенко А. А.	Каркищенко А. Н.	Одиноких Г. А.
Анциперов В. Е.	Катруца А. М.	Панов А. И.
Бахтеев О. Ю.	Копылов А. В.	Панов М. Е.
Бунакова В. Р.	Кочедыков Д. А.	Потапенко А. А.
Вальков А. С.	Кочетов Ю. А.	Пушняков А. С.
Ветров Д. П.	Красоткина О. В.	Рейер И. А.
Визильтер Ю. В.	Крымова Е. А.	Рудой Г. И.
Володин С. Е.	Кудинов М. С.	Рябенко Е. А.
Воронцов К. В.	Кузнецов М. П.	Сафонов И. В.
Гасников А. В.	Кузнецова М. В.	Сенько О. В.
Генрихов И. Е.	Кузьмин А. А.	Середин О. С.
Гнеушев А. Н.	Кулунчаков А. С.	Сотнезов Р. М.
Голиков А. И.	Кушнир О. А.	Стенина М. М.
Гончаров А. В.	Ланге М. М.	Стрижов В. В.
Гороховский К. Ю.	Ломов Н. А.	Сулимова В. В.
Двоенко С. Д.	Лукашевич Н. В.	Талипов К. И.
Дударенко М. А.	Майсурадзе А. И.	Таханов Р. С.
Дьяконов А. Г.	Максимов Ю. В.	Торшин И. Ю.
Жариков И. Н.	Матвеев И. А.	Трёкин А. Н.
Животовский Н. К.	Матросов М. П.	Турдаков Д. Ю.
Загоруйко Н. Г.	Местецкий Л. М.	Федоряка Д. С.
Зайцев А. А.	Миркин Б. Г.	Фрей А. И.
Ивахненко А. А.	Михеева А. В.	Хачай М. Ю.
Игнатов А. Д.	Мнухин В. Б.	Хританков А. С.
Игнатов Д. И.	Мотренко А. П.	Царьков С. В.
Игнатъев В. Ю.	Мурашов Д. М.	Черепанов Е. В.
Инякин А. С.	Неделько В. М.	Чичева М. А.
Ишкина Ш. Х.	Нейчев Р. Г.	Чуличков А. И.
Карасиков М. Е.	Новик В. П.	Янина А. О.

Reviewers

Aduenko A.	Kochedykov D.	Panov M.
Antsiperov V.	Kochetov Yu.	Potapenko A.
Bakhteev O.	Kopylov A.	Pushnyakov A.
Bunakova V.	Krasotkina O.	Reyer I.
Cherepanov E.	Krymova E.	Rudoy G.
Chicheva M.	Kudinov M.	Ryabenko E.
Chulichkov A.	Kulunchakov A.	Safonov I.
Dudarenko M.	Kushnir O.	Sen'ko O.
Dvoenko S.	Kuz'min A.	Seredin O.
D'yakonov A.	Kuznetsov M.	Sotnezov R.
Fedoryaka D.	Kuznetsova M.	Stenina M.
Frei A.	Lange M.	Strizhov V.
Gasnikov A.	Lomov N.	Sulimova V.
Genrikhov I.	Lukashevich N.	Takhanov R.
Gneushev A.	Maksimov Yu.	Talipov K.
Golikov A.	Matrosov M.	Torshin I.
Goncharov A.	Matveev I.	Trekin A.
Gorokhovskiy K.	Maysuradze A.	Tsar'kov S.
Ignat'ev V.	Mestetskiy L.	Turdakov D.
Ignatov A.	Mikheeva A.	Val'kov A.
Ignatov D.	Mirkin B.	Vetrov D.
Inyakin A.	Mnukhin V.	Vizil'ter Yu.
Ishkina Sh.	Motrenko A.	Volodin S.
Ivakhnenko A.	Murashov D.	Vorontsov K.
Karasikov M.	Nedel'ko V.	Yanina A.
Karkishchenko A.	Nejchev R.	Zagorujko N.
Katrutsa A.	Novik V.	Zajtsev A.
Khachay M.	Odinokikh G.	Zharikov I.
Khritankov A.	Panov A.	Zhivotovskiy N.

Краткое оглавление

Теория и методы машинного обучения	10
Дискретная оптимизация и сложность вычислений	46
Анализ сигналов и временных рядов	62
Обработка изображений	76
Анализ и распознавание изображений	94
Морфология изображений	110
Анализ текстов и информационный поиск	128
Анализ биомедицинских сигналов	144
Биоинформатика	166
Геоинформатика	174
Прикладные системы	178
Программная инженерия	188
Содержание	202
Авторский указатель	221

Brief contents

Machine Learning	10
Discrete Optimization and Computational Complexity	46
Signal and Time Series Analysis	62
Image Processing	76
Image Analysis and Recognition	94
Morphological Image Processing	110
Text Analysis and Information Retrieval	128
Biomedical Signal Analysis	144
Bioinformatics	166
Geoinformatics	174
Applied Systems	178
Software Engineering	188
Contents	202
Author index	224

Градиентные методы оптимизации гиперпараметров моделей глубокого обучения

Бахтеев Олег Юрьевич

bakhteev@phystech.edu

Москва, Московский физико-технический институт (ГУ)

Решается задача оптимизации гиперпараметров моделей глубокого обучения. *Модель* — суперпозиция функций, решающая задачу классификации или регрессии. *Гиперпараметры* модели — параметры распределения параметров модели.

Нахождение оптимальных гиперпараметров сильно влияет на правдоподобие модели. Рассматриваются методы, основанные на градиентной оптимизации. Основным их преимуществом перед другими методами оптимизации гиперпараметров является возможность эффективно оптимизировать большое количество гиперпараметров одновременно. Другой особенностью данных алгоритмов является возможность их применения для произвольной дифференцируемой функции потерь. В данной работе сравниваются градиентные методы оптимизации гиперпараметров: жадный алгоритм оптимизации, DgMAD и HOAG. Сложность рассматриваемых алгоритмов сопоставима со сложностью оптимизации параметров модели.

В экспериментальной части в качестве критерия выбора модели выступают вариационная нижняя оценка правдоподобия модели и ошибка на валидационной части выборки. Исследуется поведение алгоритмов на выборках большой мощности, таких как WISDM и MNIST. Рассматриваемые алгоритмы показывают высокое качество в случае, когда количество гиперпараметров велико. Экспериментально проиллюстрировано, что в данном случае оптимизация гиперпараметров методом случайного поиска дает значительно худшие результаты за то же количество итераций.

Исследование поддержано РФФИ, проект 16-07-01160.

- [1] Бахтеев О. Ю., Попова М. С., Стрижов В. В. Системы и средства глубокого обучения в задачах классификации // Системы и средства информатики Москва: Институт Проблем Информатики, РАН, 2016. — С. 4–22. http://www.ipiran.ru/journal_system/article/08696527160201.html.

Gradient-based hyperparameter optimization for deep learning models

Bakhteev Oleg

bakhteev@phystech.edu

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

The paper describes methods of hyperparameter optimization. A model is a function superposition, hyperparameters are parameters of parametrical distribution.

The paper describes gradient-based hyperparameter optimization methods, which allow to optimize a large amount of hyperparameters effectively. Gradient-based methods can be used with arbitrary differential loss function. The reviewed algorithms are HOAG, DrMad and greedy gradient-based algorithm.

The experiments were conducted on WISDM accelerometer records dataset, MNIST handwritten digits dataset and simulation data. The experiments show that the reviewed methods handle hyperparameter optimization effectively in case when the amount of the hyperparameters is large. We compare the reviewed methods with random search using the same amount of optimization iterations. The experiments show that the gradient-based methods can optimize hyperparameters significantly better.

This research is supported by RFBR project 16-07-01160.

- [1] *Bakhteev O., Popova M., and Strijov V.* Systems and means of deep learning for classification problems // Systems and Means of Informatics, Moscow: Institute of Informatics Problems, Russian Academy of Sciences, 2016, 2. — p. 4–22. http://www.ipiran.ru/journal_system/article/08696527160201.html.

Построение оптимального коллективного решения в кластерном анализе на основе усредненной коассоциативной матрицы и индексов качества группировки

Бериков Владимир Борисович^{1,2}

berikov@math.nsc.ru

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

²Новосибирск, Новосибирский государственный университет

В работе [1] проведено теоретическое и экспериментальное исследование метода построения кластерного ансамбля, основанного на наборе различных алгоритмов кластерного анализа. Коллективное решение строится путем анализа усредненной коассоциативной матрицы, при нахождении которой учитываются индексы качества вариантов группировки. Доказано, что элементы матрицы определяют псевдометрику. Для обоснования разработанного метода использована вероятностная модель ансамблевой попарной классификации с латентными классами. Получены аналитические зависимости между оценками качества решения и характеристиками ансамбля (числом его элементов, ожидаемым значением и дисперсией индекса качества, показателями устойчивости базовых алгоритмов). Сделаны выводы о том, что при выполнении определенных условий модели качество коллективного решения растет при увеличении числа элементов ансамбля, показателей устойчивости базовых алгоритмов, ожидаемых индексов качества группировки и уменьшении дисперсии индексов. Показано, что учет индексов особенно важен при решении задач, в которых возможно получение лишь небольшого числа элементов ансамбля (например, для больших объемов данных и ограничений на время принятия решения). В рамках модели найдено выражение для оптимальных весов, для которых минимальна оценка вероятности ошибки классификации.

Работа поддержана Министерством науки и образования РФ в рамках программы 5-100.

- [1] *Berikov V.* Construction of an optimal collective decision in cluster analysis on the basis of an averaged co-association matrix and cluster validity indices // Pattern Recognition and Image Analysis. 2017. Vol. 27(2). P. 153-165

Construction of an Optimal Collective Decision in Cluster Analysis on the Basis of an Averaged Co-Association Matrix and Cluster Validity Indices

Berikov Vladimir^{1,2}

berikov@math.nsc.ru

¹Novosibirsk, Sobolev Institute of Mathematics SB RAS

²Novosibirsk, Novosibirsk State University

In the work [1] we have carried out a theoretical and experimental investigation of a method for constructing a cluster ensemble based on a set of various cluster analysis algorithms. A collective decision is constructed by analyzing an averaged co-association matrix that is found with regard to the validity indices of the variants of grouping obtained. We have proved that the elements of the co-association matrix define a pseudometric. To substantiate the method developed, we have used a probabilistic model of ensemble pairwise classification with latent classes. Using the model, we have obtained analytic relations between the estimates for the quality of a decision and the characteristics of the ensemble (the number of its elements, the expected value and the variance of the validity index, and the stability indices of the base algorithms). We have drawn conclusions that, under certain conditions of the model, the quality of the collective decision increases as the number of elements of the ensemble, the stability indices of the base algorithms, and the expected cluster validity indices increase and the variance of the indices decreases. We have shown that the consideration of validity indices is especially important when solving problems where one can obtain only a small number of elements of the ensemble (for example, for large data size and constraints on the decision making time). Within the model, we have obtained an expression for optimal weights for which the estimate for the probability of classification error is minimal.

The work is supported by Russian Ministry of Science and Education under the 5-100 Excellence Programme.

- [1] *Berikov V.* Construction of an optimal collective decision in cluster analysis on the basis of an averaged co-association matrix and cluster validity indices // Pattern Recognition and Image Analysis. 2017. Vol. 27(2). P. 153-165

Прогноз проницаемости горной породы с помощью символьной регрессии

Бочкарёв Артём Максимович^{1*} artem.bochkarev@phystech.edu

*Софронов Иван Львович*¹ ilsofronov@gmail.com

Стрижов Вадим Викторович^{1,2} strijov@phystech.edu

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Федеральный исследовательский центр «Информатика и управление» РАН

Целью работы является построение интерпретируемых моделей физических явлений. Предлагается использовать комбинацию моделей символьной регрессии и нейронной сети с целью повышения точности прогноза при уменьшении вычислительной сложности. Метод символьной регрессии заключается в нахождении как можно более точной и простой суперпозиции признаков. Каждая суперпозиция генерируется при помощи грамматики порождения функций на основе элементарных формул и представляется в виде дерева, по которому восстанавливается конечная формула.

На первом этапе построения комбинированной модели отыскиваются несколько лучших суперпозиций исходных признаков при помощи генетического алгоритма. На втором этапе строится нейронная сеть с одним скрытым слоем, на вход которой подаются полученные суперпозиции.

Предложенная модель тестируется на четырех выборках физических измерений. В экспериментах сравнивается точность работы комбинированной модели и обычной нейросети с одним скрытым слоем. Помимо уменьшения числа нейронов скрытого слоя, требуемых для достижения одной и той же точности, наблюдается устойчивое улучшение максимальной точности при использовании комбинированной модели.

Работа поддержана грантами РФФИ № 16-07-01154 и РНФ № 15-11-00015.

[1] *Бочкарёв А. М., Софронов И. Л., Стрижов В. В.* Порождение экспертно-интерпретируемых моделей для прогноза проницаемости горной породы // Информатика и ее применения, 2017.

Permeability prediction using symbolic regression

Bochkarev Artem^{1*}

artem.bochkarev@phystech.edu

*Sofronov Ivan*¹

ilsofronov@gmail.com

Strijov Vadim^{1,2}

strijov@phystech.edu

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

²Moscow, Federal Research Center “Computer Science and Control”
of RAS

The goal is to build interpretable models of physical events. The authors propose a method of combining symbolic regression and a neural network to achieve best prediction accuracy and decrease computational costs. The symbolic regression is an approach to find a simple, yet an accurate superposition of features. Each superposition is generated using grammar of generating functions, based on elementary formulas. These superpositions are represented as trees that recover the original formula.

During the first step several optimal superpositions of original features are found, using genetic algorithm. On the second step a neural network with one hidden layer is built on top of these superpositions.

The proposed method was tested on four datasets of the physical measurements. The accuracy of combined model is compared with neural network with one hidden layer. Aside from decrease in the amount of neurons needed for the same accuracy, there is stable improvement of accuracy when using combined model.

This research is funded by RFBR, grant 16-07-01154 and RSCF, grant 15-11-00015.

[1] *Bochkarev A., Sofronov I., Strijov V.* Generation of expertly-interpreted models for prediction of core permeability // Informatics and Applications, 2017.

Структурно-функциональный анализ и синтез глубоких конволюционных нейронных сетей

Визильтер Юрий Валентинович^{1*}

viz@gosniias.ru

*Горбачевич Владимир Сергеевич*¹

gvs@gosniias.ru

¹Москва, ФГУП «Государственный научно-исследовательский институт авиационных систем» (ФГУП «ГосНИИАС»)

Предложен общий подход к Структурно-Функциональному Анализу и Синтезу (СФАС) глубоких конволюционных нейронных сетей (CNN), который позволяет регулярным образом определить: из каких типовых структурно-функциональных элементов (СФЭ) могут строиться CNN; каковы необходимые математические свойства СФЭ; какие комбинации СФЭ являются допустимыми; каковы возможные пути построения и обучения глубоких сетей для анализа и распознавания нерегулярных, неоднородных или сложно структурированных данных (таких как нерегулярные массивы, данные различного размера и различной природы, деревья, скелеты, графовые структуры, 2D, 3D и ND облака точек, триангулированные поверхности, аналитические описания данных и т.п.).

Формализм СФАС основан на рассмотрении процесса фичеризации, т.е. последовательной (от уровня к уровню CNN) «перегонки» информации о геометрии и топологии объекта в глобальный вектор признаков. Исходя из этого, в работе определен необходимый набор СФЭ и предложена многомерная систематическая таблица, каждая из «строк» и «столбцов» которой соответствует набору возможных реализаций определенных СФЭ или способов их комбинирования и использования, а каждой ячейке таблицы соответствует определенная структура CNN. Выбор общей структуры такой таблицы соответствует задаче структурно-функционального анализа CNN, а выбор конкретных ячеек для решения задач распознавания – задаче структурно-функционального синтеза CNN.

Работа выполнена при финансовой поддержке РФФИ, грант № 16-11-00082.

[1] *Визильтер Ю. В., Горбачевич В. С.* Структурно-функциональный анализ и синтез глубоких конволюционных нейронных сетей // Изв. РАН. ТиСУ, 2017 (в печати).

Structural-Functional Analysis and Synthesis of Deep Convolutional Neural Networks

*Vizilter Yury*¹★

viz@gosniias.ru

*Gorbatsevich Vladimir*¹

gvs@gosniias.ru

¹Moscow, The Federal State Unitary Enterprise “State Research Institute of Aviation Systems” (FGUP “GosNIIAS”)

The generalized approach to structural-functional analysis and synthesis (SFAS) of deep convolutional neural networks (CNN) is proposed. This approach proclaims a regular way to determine the set of some structural-functional elements (SFE), which are the unified building blocks for generalized CNN architecture construction. SFAS also explores the required mathematical properties of SFEs and their possible combinations. In result, SFAS provides a technique for forming and learning CNNs, which could process and recognize the irregular, non-homogeneous and complex structured data, such as irregular arrays, size-variable data, fused data of different types or nature, tree-like structures, skeletons, graphs and nets, 2D, 3D and ND point clouds, triangulated meshes, analytical surfaces, etc.

The proposed SFAS formulation is based on analysis of step-by-step (CNN level-by-level) “featurization” process, which transforms the spatially distributed input geometric and topological information to the output global feature vector. The set of SFEs required for implementation of such featurization is extracted from the typical CNN structure. Then the SFAS “systematic table” is proposed for supporting the CNN structure analysis and synthesis. It contains SFEs as “columns” and different data structures and types as “rows”. So, the “cells” of this SFAS table should contain the concrete implementation of CNN architectural and functional elements, which allows processing and recognizing the input patterns of given structure type.

This research is funded by RSF, grant 16-11-00082.

- [1] *Vizilter Yu. and Gorbatsevich V.* Structural-Functional Analysis and Synthesis of Deep Convolutional Neural Networks // Computer and Systems Sciences International, 2017 (in press).

О полных решающих деревьях в задаче восстановления регрессии

*Генрихов Игорь Евгеньевич*¹

ingvar1485@rambler.ru

*Дюкова Елена Всеволодовна*²

edjukova@mail.ru

Журавлёв Вадим Игоревич^{3*}

vadim091294@gmail.com

¹Химки, ООО «Мобайл парк ИТ»

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

³Москва, Московский государственный университет
им. М. В. Ломоносова

Рассматривается одна из актуальных задач машинного обучения – задача восстановления регрессии. Среди существующих подходов к решению этой задачи выделяют подход, основанный на построении регрессионных решающих деревьев (РРД). Наиболее известные алгоритмы синтеза регрессионных деревьев (например, алгоритмы CART и Random Forest) основаны на использовании бинарных РРД. Реже используются k -арные РРД.

В [1] рассматриваемая задача решается на основе построения так называемых полных РРД. По сравнению с классическим РРД конструкция полного РРД позволяет более существенно использовать имеющуюся информацию, поскольку на каждой итерации в ветвлении участвуют все признаки, удовлетворяющие выбранному критерию. При этом описание распознаваемого объекта может порождаться несколькими ветвями РРД. Каждая такая ветвь участвует в процедуре голосования. Ранее подход к построению полных РРД был исследован авторами на задачах с целочисленными данными и показал повышение качества решения по сравнению с рядом других методов синтеза регрессионных деревьев. Получены новые результаты, связанные с построением полных k -арных РРД для случая вещественнозначной информации, который, как известно, наиболее часто встречается на практике.

Работа поддержана грантом РФФИ № 16-01-00445.

- [1] *Генрихов И. Е., Дюкова Е. В., Журавлёв В. И.* Построение и исследование полных решающих деревьев для задачи восстановления регрессии в случае вещественнозначной информации // Машинное обучение и анализ данных, 2017.

On full decision trees in regression restoration problem

*Genrikhov Igor*¹

ingvar1485@rambler.ru

*Djukova Elena*²

edjukova@mail.ru

Zhuravlyov Vadim^{3*}

vadim091294@gmail.com

¹Himki, LLC Mobail park IT

²Moscow, Dorodnicyn Computing Centre of FRC CSC RAS

³Moscow, Lomonosov Moscow State University

Regression restoration problem as the one of the actual problems of machine learning is considered. The approach based on creation of the regression decision trees (RDT) is highlighted among the existing approaches to solve this problem. The most known algorithms of the regression trees synthesis (e.g., algorithms CART and Random Forest) are based on use of binary RDT. RDT with k -ary structure are rarely used.

The considered problem is solved based on the creation of the so-called full RDT in [1]. In comparison with classical RDT the design of full RDT allows use of the available information more efficient, as all features meeting the chosen criterion are involved in the branching at each iteration. Herewith, several branches of RDT can generate the description of the recognizable object. Each such branch participates in the voting procedure. Earlier, the approach to the construction of the full RDT was investigated by the authors based on problems with integer data and has shown improvement of the quality of the decision in comparison with other methods of synthesis of regression trees. The new results related to the creation of the full k -ary RDT for a case of real-valued information, which, as we know, is most often encountered in practice, are received.

This research is supported by RFBR, grant 16-01-00445.

- [1] *Genrikhov I. E., Djukova E. V., Zhuravlyov V. I.* Construction and investigation of full regression trees in regression restoration problem in the case of real-valued information // Journal of Machine Learning and Data Analysis, 2017.

Построение интерпретируемых моделей глубокого обучения в задаче социального ранжирования

Гончаров Алексей Владимирович^{1*}

alex.goncharov@phystech.edu

Стрижов Вадим Викторович^{1,2}

strijov@phystech.edu

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Федеральный исследовательский центр «Информатика и управление» РАН

В работе предложена процедура построения интерпретируемых скоринговых моделей методами глубокого обучения. В задаче кредитного скоринга определяется уровень кредитоспособности заемщика. Для этого используется анкета заемщика, содержащая признаки как в линейных, так и в ординальных и номинальных шкалах.

Цель работы — создать модель глубокого обучения для построения скоринговой карты как суперпозицию процедур порождения признаков. Необходимо при этом удовлетворить требования к интерпретируемости и качеству модели. Для решения задачи используется набор эксплуатационных критериев качества в единой постановке задачи оптимизации.

Процедура включает все шаги, которые аналитик выполняет при построении эксплуатируемой модели. А именно, фильтрация выбросов и заполнение пропусков в данных, процедуры группировки и сегментации, создание интерпретируемых комбинаций признаков, снижение размерности признакового пространства, построение моделей классификации и ранжирования и вычисление критериев качества. Комбинация этих шагов в единой процедуре решения сводит задачу оптимизации к частично-целочисленной.

Данная работа выполнена при финансовой поддержке Российской Федерации (Соглашение № 05.Y09.21.0018), поддержана грантом РФФИ № 16-07-01163

- [1] *Гончаров А. В.* Continuous Time Series Alignment in Human Actions Recognition // Proceedings of the AINL FRUCT 2016, Санкт-Петербург, 2016. <https://fruct.org/publications/abstract-AINL-FRUCT-2016/>.

Deep interpreted models in social ranking task

*Goncharov Alexey*¹*

alex.goncharov@phystech.edu

Strijov Vadim^{1,2}

strijov@phystech.edu

¹Moscow, Moscow Institute of Physics and Technology

²Moscow, Federal Research Center “Computer Science and Control”
of RAS

The research develops a complex procedure for constructing interpreted scoring models using deep methods. The problem of credit scoring determines the level of creditworthiness of the borrower. A borrower’s questionnaire is used. It contains both numerical (age, income) and categorical features (gender, profession).

The purpose is create a deep model to build a scorecard as a procedure superposition. It is necessary to keep to the interpretability and quality requirements. A set of operational quality criteria is used in the united optimization problem.

The procedure includes all steps, which an analyst performs to build a model: outliers filtering, filling gaps, grouping, segmentation procedures, creating interpretable feature combinations, reducing the dimensionality of the feature space, building classification ranking models, and calculating quality criteria. The combination of these steps in a end-to-end decision procedure reduces the optimization problem to the mixed integer one.

This research was supported by Government of the Russian Federation (Agreement 05.Y09.21.0018), funded by RFBR, grant 16-07-01163.

- [1] *Goncharov A. V.* Continuous Time Series Alignment in Human Actions Recognition // Proceedings of the AINL FRUCT 2016, St. Petersburg, 2016. <https://fruct.org/publications/abstract-AINL-FRUCT-2016/>.

Исследование обусловленности матриц парных сравнений при их метрической коррекции

Двоенко Сергей Данилович^{1*}

sergedv@yandex.ru

*Пшеничный Денис Олегович*¹

denispshenichny@yandex.ru

¹Тула, Тульский государственный университет

В интеллектуальном анализе данных результаты исследований часто представлены в виде парных сравнений сходства или различия объектов. Для корректного применения алгоритмов машинного обучения результаты парных сравнений необходимо погрузить в метрическое пространство. Одним из условий корректного погружения является неотрицательная определенность матрицы парных близостей элементов множества друг с другом. В этом случае неотрицательные близости интерпретируются как скалярные произведения векторов в положительном квадранте гипотетического признакового пространства, а соответствующие различия представляют собой расстояния.

На практике применяют различные способы оценки сходства или различия элементов множества. Во многих случаях такие функции сравнения не обладают свойствами функций близостей или расстояний. Поэтому возникает необходимость в метрической коррекции произвольных экспериментальных матриц парных сравнений для обеспечения положительной определенности соответствующих им нормированных матриц скалярных произведений.

Но естественное требование минимизации отклонений значений скорректированных матриц от их исходных значений обычно приводит к плохо обусловленным матрицам скалярных произведений с большим числом обусловленности.

В данной работе исследуется обусловленность скорректированных матриц скалярных произведений и предлагается подход к уменьшению числа обусловленности.

Работа частично поддержана грантами РФФИ № 17-07-00319, № 17-07-00993, № 17-07-00436.

- [1] *Двоенко С. Д., Пшеничный Д. О.* Обусловленность матриц парных сравнений при коррекции метрических нарушений // Машинное обучение и анализ данных, 2017.

Investigation of matrix conditionality under metric correction of pairwise comparisons

Dvoenko Sergey^{1*}

sergedv@yandex.ru

*Pshenichny Denis*¹

denispshenichny@yandex.ru

¹Tula, Tula State University

In modern intelligent data analysis, results of investigations are usually represented by mutual pairwise comparisons of similarity or dissimilarity of objects. It needs to immerse results of pairwise comparisons into some metric space for correct using of machine learning algorithms. One of conditions of the correct immersion of set elements is the non-negative definite matrix of their pairwise similarities. In this case, non-negative similarities represent scalar products of vectors in the positive quadrant of an imaginary feature space, while corresponding dissimilarities represent distances.

Various similarity and dissimilarity measurements are used in practice. Nevertheless, not all of them are correct as metric functions. Therefore, it needs to use metric corrections of real experimental matrices of pairwise comparisons to support the positive definiteness of corresponding matrices of standard scalar products.

Unfortunately, the natural limit to minimize deviations of corrected values from initial ones leads to ill-conditioned matrices of scalar products with the large condition number.

In this paper, we investigate a way to improve the conditionality of matrices of pairwise comparisons. This research is funded by RFBR, grants 17-07-00319, 17-07-00993, 17-07-00436.

[1] *Dvoenko S., Pshenichny D.* The conditionality of matrices of pairwise comparisons after metric corrections // Journal of Machine Learning and Data Analysis, 2017.

О дуализации над произведением цепей

*Дюкова Елена Всеволодовна*¹

edjukova@mail.ru

Масляков Глеб Олегович^{2*}

gleb-mas@mail.ru

*Прокофьев Пётр Александрович*¹

p_prok@mail.ru

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

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

Пусть $P = P_1 \times \dots \times P_n$, где P_1, \dots, P_n — конечные частично упорядоченные множества. Считается, что элемент $y = (y_1, \dots, y_n) \in P$ следует за $x = (x_1, \dots, x_n) \in P$, если y_i следует за x_i при $i = 1, \dots, n$. Пусть $R \subset P$ и $R^+ = \{x \in P \mid \exists a \in R, a \preceq x\}$. Задача построения двойственного к R множества $I(R)$, состоящего из элементов $a \in P \setminus R^+$ таких, что для любого $x \in P \setminus R^+$ такого, что $x \neq a$, отношение $a \preceq x$ не выполняется (т. е. $I(R)$ состоит из максимальных элементов множества $P \setminus R^+$), называется дуализацией над произведением частичных порядков. Данная задача является одной из центральных труднорешаемых задач дискретной математики.

Практически важным случаем является случай, когда каждое P_i является цепью, т.е. любые два элемента в P_i сравнимы. Если при этом $|P_i| = 2$, то поставленная задача эквивалентна поиску неприводимых покрытий булевой матрицы размера $m \times n$, где m — число элементов в R , и возникает, в частности, при конструировании логических процедур классификации по прецедентам. В [1] рассматривается более общий случай, а именно, когда каждое P_i является цепью и $|P_i| = k, k \geq 2$. Показывается, что задача сводится к поиску некоторого подмножества множества неприводимых покрытий булевой матрицы размера $m \times kn$. Приводятся результаты численных экспериментов, базирующиеся на асимптотически оптимальном перечислении неприводимых покрытий.

Работа поддержана грантом РФФИ № 16-01-00445.

[1] Дюкова Е. В., Масляков Г. О., Прокофьев П. А. О дуализации над произведением цепей // Машинное обучение и анализ данных, 2017.

About dualization over product of chains

*Djukova Elena*¹

edjukova@mail.ru

Maslyakov Gleb^{2*}

gleb-mas@mail.ru

*Prokofyev Petr*¹

p_prok@mail.ru

¹Moscow, Federal Research Center “Computer Science and Control” of RAS

²Moscow, Lomonosov Moscow State University

Let $P = P_1 \times \dots \times P_n$, where P_1, \dots, P_n are finite partially ordered sets. Assume that $y = (y_1, \dots, y_n) \in P$ follows $x = (x_1, \dots, x_n) \in P$, if y_i follows x_i for $i = 1, 2, \dots, n$. To denote that $y \in P$ follows $x \in P$ and $y \neq x$, the notation $x \prec y$ is used later. Let $R \subset P$, $R^+ = R \cup \{x \in P \mid \exists a \in R, a \prec x\}$. The problem of constructing the dual to R set $I(R)$ consisting of elements $a \in P \setminus R^+$ such that for any $x \in P \setminus R^+$, $x \neq a$, the relation $a \prec x$ does not hold, is called a dualization over the product of partially ordered sets. This problem is one of the central intractable problems of discrete mathematics.

A practically important case is the case where each P_i is a chain, i.e. any two elements of P_i are comparable. If, in addition, $|P_i| = 2$, then the problem posed is equivalent to the search for irreducible coverings of a Boolean matrix of size $m \times n$, where m is the number of elements in R , and arises, in particular, in the construction of logical classification procedures by precedents. In [1], a more general case is considered, namely, when each P_i is a chain and $|P_i| = k$, $k \geq 2$. It is shown that the problem reduces to the search for a subset of the set of irreducible coverings of a Boolean matrix of size $m \times n$. The results of numerical experiments based on an asymptotically optimal enumeration of irreducible coverings are presented.

The work is supported by the RFBR, grant 16-01-00445.

- [1] *Djukova E. V., Maslyakov G. O., Prokofyev P. A.* About dualization over product of partially ordered sets // Journal of Machine Learning and Data Analysis, 2017.

Локальные модели для классификации объектов сложной структуры

Исаченко Роман Владимирович^{1,2} isa-ro@yandex.ru

Жариков Илья Николаевич^{1,2} ilya250894@gmail.com

Бочкарёв Артём Максимович^{1,2}* artem.bochkarev@phystech.edu

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Сколковский институт науки и технологий

В работе рассматривается задача классификации объектов без явного признакового представления. Данные содержат временные ряды ускорения, полученные с акселерометра смартфона. Задача состоит в предсказании вида физической активности человека по временному ряду. Временной ряд представляет собой объект сложной структуры без явного признакового описания. В работе предлагается подход к генерации признаков временных рядов, рассматривая их как объекты сложной структуры. В качестве признакового описания используются параметры локальных моделей. Сгенерированные признаки позволяют получить приемлемое качество классификации и требуют умеренных вычислительных ресурсов.

Проблема классификации объектов сложной структуры разбивается на две несвязанные процедуры. Первая извлекает информативные признаки. Вторая классифицирует объекты, используя порожденные признаки.

Данная работа нацелена на сравнение различных методов генерации признаков: экспертные функции, авторегрессионная модель и анализ сингулярного спектра. Авторы предлагают новый метод порождения признаков. Сегменты временного ряда аппроксимируются кубическими сплайнами. Сплайны генерируют гладкую кривую с достаточным качеством аппроксимации.

Эксперимент проводился на двух датасетах с акселерометра: WISDM, USC-HAD. Сравнивались качества упомянутых методов извлечения признаков и различных моделей классификации.

Работа поддержана грантом РФФИ № 16-07-01154.

- [1] *Исаченко Р. В., Жариков И. Н., Бочкарёв А. М.* Локальные модели для классификации объектов сложной структуры // Journal of Machine Learning and Data Analysis, 2016. — №. 13. <http://jmla.org/papers/doc/2016/no1/Isachenko2016MetricLearning.pdf>.

Local models for classification of complex structured objects

Isachenko Roman^{1,2}

isa-ro@yandex.ru

Zharikov Ilya^{1,2}

ilya250894@gmail.com

Bochkarev Artem^{1,2}★

artem.bochkarev@phystech.edu

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

²Moscow, Skolkovo Institute of Science and Technology

The article investigates a problem of multiclass classification of objects without explicit feature set. The data which were used are time series from smartphone accelerometer. The problem is to predict human activity, using acceleration time series. In this problem statement, time series is a complex structured object without explicit feature set. The paper proposes method of feature generation for complex structured objects. The features will correspond to parameters of local approximation models. Features generated in such way allow decent classification quality and require low computational resources.

The problem of classification of complex structured objects is divided in two independent procedures. First, informative features are extracted and then the classification model is built on these features.

This paper goal is to compare different approaches to feature generation: expert functions, autoregressive model and singular spectrum analysis. The new method of feature generation is proposed. This segments of time series are approximated with cubic splines, which generate smooth curve with sufficient quality of approximation.

The experiment was conducted on two real datasets: WISDM and USC-HAD. The experiment shows comparison of different feature generation procedures and classification models.

This research is funded by RFBR, grant 16-07-01154.

- [1] *Isachenko R., Zharikov I., Bochkarev A.* Local models for classification of complex structured objects // Journal of Machine Learning and Data Analysis, 2016. — No 13. <http://jmla.org/papers/doc/2016/no1/Isachenko2016MetricLearning.pdf>.

Методы оценивания гиперпараметров в задачах динамического анализа стиля инвестиций

*Красоткина Ольга Вячеславовна*¹ o.v.krasotkina@yandex.ru
Моттль Вадим Вячеславович^{2*} vmottl@yandex.ru
*Марков Михаил*³ mmarkov@markovprocesses.com

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

²Москва, Федеральный исследовательский центр «Информатика и управление» РАН

³Summit, NJ, USA, Markov Processes International Inc.

Проблема оценивания нестационарной регрессии неизбежно связана с необходимостью выбора волатильности регрессионной модели — от полной стационарности мгновенных моделей до их полной независимости. В стационарном случае число оцениваемых коэффициентов регрессии равно числу регрессоров, а при отсутствии гладкости коэффициент увеличения размерности оцениваемого вектора равен длине временного ряда. Мы рассматриваем семейство непрерывно вложенных априорных распределений вероятностей, определяемое спецификой модели нестационарных данных. Размерность полного вектора параметров фиксирована, но свобода выбора его значения «мягко» ограничена семейством вложенных априорных распределений, которое содержит несколько гиперпараметров. Данный доклад преследует три цели. Во-первых, согласно специфике нестационарной регрессии, мы модифицируем три общепринятых метода оценивания гиперпараметров в моделях данных, а именно, методы кросс-валидации по принципу скользящего контроля, максимизации маргинального распределения данных (Evidence Maximization) и гипотетической кросс-валидации. Во-вторых, мы экспериментально сравниваем эти методы как на модельных, так и на реальных данных. В третьих, на основе разработанной методологии мы предлагаем новый подход к проблеме оценивания скрытой динамики инвестиционного портфеля по доступным рыночным факторам.

Работа поддержана грантом РФФИ № 17-07-00993.

- [1] *Krasotkina O., Mottl V., Markow M., et al.* Methods of hyperparameter estimation in time-varying regression models with application to dynamic style analysis of investment portfolios // Lecture Notes in Computer Science, Vol. 10358, Springer, 2017, pp. 431-450.

Methods of hyperparameter estimation in time-varying regression models with application to dynamic style analysis of investment portfolios

*Krasotkina Olga*¹

`o.v.krasotkina@yandex.ru`

Mottl Vadim^{2*}

`vmottl@yandex.ru`

*Markov Michael*³

`mmarkov@markovprocesses.com`

¹Moscow, Lomonosov Moscow State University

²Moscow, Federal Research Center “Computer Science and Control” of RAS

³Summit, NJ, USA, Markov Processes International Inc.

The problem of estimating time-varying regression inevitably concerns the necessity to choose the appropriate level of model volatility – ranging from the full stationarity of instant regression models to their absolute independence. In the stationary case the number of regression coefficients, constituting the model parameter to be estimated, equals that of regressors, whereas the absence of any smoothness assumptions augments the dimension of the unknown vector by the factor of the time-series length. We consider here a family of continuously nested a priori probability distributions matching the specificity of time-varying data models, in which the dimension of the parameter is fixed, but the freedom of its values is softly constrained by a family of continuously nested a priori probability distributions, which contains a number of hyperparameters. The aim of this paper is threefold. First, in accordance with the specificity of the time-varying regression, we modify three commonly adopted methods of estimating hyperparameters in data models, namely, Leave-One-Out Cross Validation, Evidence Maximization and Hypothetical Cross Validation. Second, we experimentally compare these methods on both simulated and real-world data. Third, on the basis of the proposed technique we develop a new approach to the problem of detecting the hidden dynamics of an investment portfolio in respect to certain market factors.

This research is funded by RFBR, grant 17-07-00993.

- [1] *Krasotkina O., Mottl V., Markov M., et al.* Methods of hyperparameter estimation in time-varying regression models with application to dynamic style analysis of investment portfolios // Lecture Notes in Computer Science, Vol. 10358, Springer, 2017, pp. 431-450.

Быстрые последовательные методы обучения обобщенных линейных моделей зависимостей

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

*Моттль Вадим Вячеславович*¹ vmottl@yandex.ru

*Красоткина Ольга Вячеславовна*² o.v.krasotkina@yandex.ru

¹Москва, Московский физико-технический институт (ГУ)

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

Рассматривается новый принцип построения последовательных процедур восстановления произвольных зависимостей (регрессионных моделей, решающих правил распознавания образов, моделей выживаемости, порядковых моделей) по очень большим массивам прецедентов на основе обобщенного линейного подхода к проблеме обучения. Процесс обучения построен как последовательный пересчет функций Беллмана в некоторой специальной процедуре динамического программирования. Строго оптимальный результат обучения произвольной (не обязательно линейной) модели зависимости на каждом шаге относительно уже полученной обучающей совокупности сводится к минимизации очередной функции Беллмана. Показано, что для весьма широкого класса моделей допустима квадратичная аппроксимация каждой функции Беллмана в окрестности точки ее минимума, приводящая к классическому фильтру Калмана. Предлагается компенсировать неизбежную потерю оптимальности обучения дополнительной коррекцией каждого шага по классическому принципу стохастической аппроксимации, который гарантирует сходимость обучения «почти на верное». Хотя такое «дотягивание» обучения является очень медленным, что характерно для стохастической аппроксимации, оценка на каждом шаге остается в небольшой окрестности строгого минимума «глобального» критерия для обработанной части массива обучающих данных.

Работа поддержана грантом РФФИ № 17-07-00993.

- [1] *Turkov P., Krasotkina O., Mottl V., Sychugov A.* Feature selection for handling concept drift in the data stream classification // Machine Learning and Data Mining in Pattern Recognition. Lecture Notes in Computer Science, 2016, Vol. 9729, pp. 614-629.

Algorithms for dynamic style analysis of investments from a large number of stock market indicies

Malenichev Anton^{1*}

malenichev@mail.ru

*Mottl Vadim*¹

vmottl@yandex.ru

*Krasotkina Olga*²

o.v.krasotkina@yandex.ru

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

²Moscow, Lomonosov Moscow State University

We consider a new method of constructing on-line procedures for estimation of arbitrary dependences (regression models, pattern-recognition rules, survival models, ranking models) from very large data sets on the basis of the generalized linear approach to the problem of training. The training process is built as iterative recomputing of Bellman functions in a special dynamic programming procedure. The strictly optimal result of training the model of the respective dependence, not obligatory a linear one, with respect to the part of the training set, which is already registered by the moment of current observation, is nothing else than the minimum point of the last Bellman function. It is shown that a quite broad class of models allow for quadratic approximation of each Bellman function in a vicinity of its minimum point, what results in the classical Kalman filter. We propose to compensate the inevitable loss of optimality of training by an additional correction of the respective step in accordance with the classical stochastic approximation principle, which guaranties the convergence of the training process “almost for sure”. Despite the fact that such a fine-tuning is extremely slow, as it is characteristic for stochastic approximation, the current estimate remains in a small vicinity of the strict minimum of the “global” criterion for the already processed part of the entire training data set.

This research is funded by RFBR, grant 17-07-00993.

- [1] *Turkov P., Krasotkina O., Mottl V., Sychugov A.* Feature selection for handling concept drift in the data stream classification // Machine Learning and Data Mining in Pattern Recognition. Lecture Notes in Computer Science, 2016, Vol. 9729, pp. 614-629.

Научное наследие Эммануила Марковича Бравермана в методологии машинного обучения

*Моттль Вадим Вячеславович*¹

vmottl@yandex.ru

Середин Олег Сергеевич^{2*}

oseredin@yandex.ru

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

²Тула, Тульский государственный университет

Гипотеза компактности, сформулированная Э. М. Браверманом в его кандидатской диссертации в 1961 году, остается главным принципом, на котором базируется вся методология машинного обучения по прецедентам. Согласно идее Бравермана достаточно, чтобы наблюдатель, вооруженный компьютером, выбрал двухместную действительную симметричную функцию на множестве объектов реального мира, измеряющую попарное несходство объектов и представляющуюся ему естественной. Такой выбор удачен, если выполняется гипотеза компактности – пары объектов, близкие в терминах выбранной функции сравнения, обладают, как правило, и близкими значениями целевой характеристики. Другим фундаментальным понятием, введенным Браверманом, является потенциальная функция, порождаемая выбранной мерой парного несходства. Потенциальная функция позволяет мысленно погрузить множество реальных объектов в гипотетическое линейное пространство, в котором она играет роль скалярного произведения. Такое линейное пространство названо Браверманом спрямляющим, поскольку всякое правило предсказания целевой характеристики, линейное относительно некоторой функции парного несходства объектов, оказывается нелинейным относительно другой функции несходства. В докладе показано, что практически все существующие методы восстановления зависимостей, в сущности, реализуют такую обобщенную линейную методологию Бравермана, различаясь лишь выбором класса потенциальных функций и способа измерения несходства. С этой точки зрения рассмотрены способы селективного комбинирования множества альтернативных функций парного несходства.

Работа поддержана грантом РФФИ № 17-07-00436.

- [1] *Mottl V., Seredin O.* Compactness hypothesis and potential functions in Machine Learning // Emmanuel Braverman's Legacy, Springer, 2017, to appear.

The scientific heritage of Emmanuel Braverman in the methodology of Machine Learning

*Mottl Vadim*¹

vmottl@yandex.ru

*Seredin Oleg*²*

oseredin@yandex.ru

¹Moscow, Federal Research Center “Computer Science and Control”
of RAS

²Tula, Tula State University

The compactness hypothesis formulated by Emmanuel Braverman in his PhD Thesis in 1961 remains the main principle underlying the entire methodology of precedent based Machine Learning. In accordance with Braverman’s idea, it is enough that the observer armed by a computer defines a two-argument real-valued function on the set of real-world entities, which would evaluate the pair-wise dissimilarity between entities and be natural from his/her viewpoint. This choice is fortunate if the compactness hypothesis is met – pairs of entities close to each other in terms of the chosen comparison function have also, as a rule, close values of the goal characteristic. Another fundamental notion introduced by Braverman is potential function produced by the chosen measure of pair-wise dissimilarity. Any potential function allows for mantally embedding the set of real-world entities into a hypothetical linear space, in which it plays the role of inner product. Such a linear space was named by Braverman the rectifying one, because any decision rule on the goal characteristic, which is linear relative to a dissimilarity function between entities, will be nonlinear with respect to another dissimilarity function. We show that practically all existing methods of dependence estimation actually imply Braverman’s generalized linear framework, and differ from each other only by the choice of the class of potential functions or the way of evaluating dissimilarity. From this point of view, we consider possible ways of selective combining a variety of alternative pair-wise dissimilarity functions.

This research is funded by RFBR, grant 17-07-00436.

- [1] *Mottl V., Seredin O.* Compactness hypothesis and potential functions in Machine Learning // Emmanuel Braverman’s Legacy, Springer, 2017, to appear.

Оценивание значимости переменных для ранговой регрессии

Неделько Виктор Михайлович

nedelko@math.nsc.ru

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

Существует большое число подходов к оцениванию значимости переменных в задачах построения решающих функций, которые применимы к задаче регрессионного анализа. Одним из таких подходов является анализ (в том числе визуальный) ROC-кривой (кривой ошибок).

Изначально понятие ROC-кривой было введено для задачи классификации (разpoznавания образов). В этом случае ROC-кривая — это по сути эмпирическая функция распределения второго класса, построенная как функция от значений эмпирической функции распределения первого класса для переменной, упорядочивающей объекты по оцененной вероятности их принадлежности первому классу.

Известны обобщения понятия ROC-кривой для регрессионного анализа (REC-кривая и RROC-кривая). Однако эти обобщения требуют явного задания прогнозируемых значений целевой переменной, в то время как для построения ROC-кривой в задаче классификации достаточно лишь упорядочивания объектов.

В работе предложены некоторые естественные обобщения понятия ROC-кривой на случай регрессионного анализа, которые более полно воспроизводят свойства ROC-кривой по сравнению с известными обобщениями.

Наиболее важным из воспроизводимых свойств является то, что предложенные кривые при случайном прогнозе приближаются к прямой, а отклонения от прямой позволяют оценить информативность «объясняющей» переменной.

Кроме того предложенные варианты ROC-кривой оказались близки конструкции эмпирического моста, что свидетельствует об их содержательности.

- [1] *Неделько В. М.* Оценивание значимости переменных в моделях ранговой регрессии // Машинное обучение и анализ данных, 2017.

Estimation of feature importance for quantile regression

Nedel'ko Victor

nedelko@math.nsc.ru

Novosibirsk, Sobolev Institute of Mathematics SB RAS

There are a large number of approaches to estimating the significance of variables in problems of constructing decision functions. One of the most important approaches is based on of the ROC curve (error curve).

Initially, the ROC curves was introduced for classification models. The extension of ROC curves for regression problems has also been investigated. Notable examples are the so-called regression error characteristic (REC) Curves and the Regression ROC (RROC) curves. However, these generalizations require the explicit specification of the predicted values of the target variable, while for constructing the ROC-curve in the classification problem one needs only the ordering of objects. There are also some other differences in essential properties of such regression ROC curves and classification ROC curves.

We propose some natural generalizations of the concept of the ROC curve for regression analysis, which more fully reproduce the properties of the ROC curve as compared to known extensions.

The most important of these properties is that the ROC curves move to a straight line, when built on random prediction. The deviations from the line allow us to estimate the importance of a variable.

The proposed variants of the ROC curve for regression were found to be close to the construction of the empirical bridge.

- [1] *Nedel'ko V.* Estimation of feature importance for quantile regression // Journal of Machine Learning and Data Analysis, 2017.

Построение модели оптимальной сложности в виде смеси экспертов

Нейчев Радослав Георгиев^{1,2*}

neychev@phystech.edu

Стрижов Вадим Викторович^{1,2}

strrijoy@phystech.edu

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Федеральный исследовательский центр «Информатика и управление» РАН

Рассматривается проблема поиска модели оптимальной сложности. Под оптимальной понимается модель, усложнение которой не приведет к значимым изменениям качества описания данных. В качестве данных рассматривается большое число взаимосвязанных временных рядов. Их источником служат датчики, производящие мониторинг показателей различных устройств. Предполагается, что пространство параметров описания временных рядов неоднородно, выборка не является простой.

Проблема поиска оптимальной сложности рассматривается применительно к двум задачам: прогнозирования (регрессии) и классификации временных рядов. Для получения точного и эффективного решения предлагается использовать смеси моделей-экспертов. Уверенность экспертов оценивается как на множестве объектов обучающей выборки, так и на множестве признаков, их описывающих. В качестве моделей-экспертов используются линейные модели и неглубокие решающие деревья. Для объединения моделей-экспертов в смесь используется нейронная сеть. В ходе вычислительного эксперимента сравниваются результаты, полученные тремя способами построения ансамблей моделей: смесь экспертов, градиентный бустинг, случайный решающий лес. Эксперимент выполнен на реальных данных, содержащих информацию о потреблении электроэнергии и погодных условиях в Польше, а так же данных, полученных с акселерометров.

Работа поддержана грантом РФФИ № 16-07-01163.

- [1] *Нейчев Р. Г.* Mixture of Experts approach to model of optimal complexity construction // International Journal of Applied and Computational Mathematics, Springer, 2017 (подана в журнал).

Mixture of Experts approach to model of optimal complexity construction

Neychev Radoslav^{1,2*}

neychev@phystech.edu

Strijov Vadim^{1,2}

strijov@phystech.edu

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

²Moscow, Federal Research Center “Computer Science and Control”
of RAS

This paper presents an approach to construction of the model of optimal complexity. Model has optimal complexity if making it more complex does not gain any valuable quality. The approach is tested against data, consisting of multiple intercorrelated time series that can be generated by different sensors. Data is not *i.i.d* and it's feature space has complex structure.

The presented approach is applied to two classical ML problems: regression and classification of time series. To obtain robust and high quality solution, a Mixture of Experts ensembling method is applied. Linear models and decision trees are used as expert-models. Neural network is used as a gating function to combine the predictions of all experts. Proposed method has been tested within computational experiment where it was compared to other ensembling methods, such as gradient boosting and Random Forest. The experiment is conducted on two real world data sets: accelerometers data and data containing formation about electricity consumption and weather conditions in Poland. This research is funded by RFBR, grant 16-07-01163.

- [1] *Neychev R.* Mixture of Experts approach to model of optimal complexity construction // International Journal of Applied and Computational Mathematics, Springer, 2017 (submitted to journal).

Сокращение размерности признакового пространства на основе критерия минимального расстояния между выпуклыми оболочками классов

Немирко Анатолий Павлович

apn-bs@yandex.ru

Санкт-Петербург, Санкт-Петербургский государственный
электротехнический университет «ЛЭТИ»

В работе рассмотрена задача сокращения размерности признакового пространства при линейном анализе 2 классов в многомерном признаковом пространстве. Решается следующая задача: в результате линейного преобразования координат нужно найти такие две ортогональные оси, при проектировании классов на которые классы максимально удалены друг от друга по каждой координате. Статистические методы визуализации не дают исчерпывающей картины области пересечений классов и не обеспечивают отображение случаев полной разделимости и случайных выбросов. В данной работе близость классов друг к другу оценивается на основе критерия минимального расстояния между их выпуклыми оболочками. Исследована возможность применения метода опорных векторов (SVM) и метода Козинца-Шлезингера (S-K) для получения ортогональных векторов редуцированного пространства. Эти методы для непересекающихся классов эквивалентны определению минимального расстояния между выпуклыми оболочками классов. Для преодоления проблемы пересекающихся классов использовано понятие глубины проникновения и даны средства для ее уменьшения. Рассмотрены алгоритмы с уменьшением (RCH), сжатием (SCH) и сдвигом выпуклых оболочек. Экспериментальные исследования посвящены применению исследованных методов сокращения размерности для визуализации биомедицинских данных в двумерном пространстве.

Работа поддержана грантами РФФИ № 15-07-01790, 16-01-00159.

- [1] *Nemirko A.* Computer geometry algorithms in feature space dimension reduction problems // *Pattern Recognition and Image Analysis*, 2017, 27 (3). — p. 387–394.

Reducing the dimension of feature space based on the criterion of minimum distance between the convex hulls

Nemirko Anatolii

apn-bs@yandex.ru

Saint Petersburg, Saint Petersburg Electrotechnical University “LETI”

One of the goals of multidimensional feature space dimension reduction is to visualize the arrangement of classes on a plane. The quality of such a representation depends on the reduction method and the measure of the closeness of classes. In this paper, we propose a technology of dimension reduction to two dimensions based on the calculation of convex hulls of classes and determination of their intersection degree. Statistical visualization methods do not give an exhaustive picture of the region of classes intersection and do not provide graphic representation of total separability and outliers. In this work, the proximity of classes to each other is based on the criterion of minimum distance between their convex hulls. The possibility of using support vector machine method (SVM) and Kozinets-Schlesinger (S-K) method to obtain the orthogonal vectors in the reduced space are investigated. These methods for disjoint classes are equivalent to specifying the minimum distance between the convex hulls of classes. To overcome the problem of intersecting classes the concept of penetration depth and the means for its reduction are used. The algorithms with the reduction (RCH), compression (SCH) and a shift of convex hulls are considered. Experiments carried out show that this technology on the training set makes it possible to identify the minimum potentially achievable intersection degree of classes intersecting in a multidimensional space and to obtain a correct mutual mapping of classes on the plane

This research is funded by RFBR, grants 15-07-01790, 16-01-00159.

- [1] *Nemirko A.* Computer geometry algorithms in feature space dimension reduction problems // Pattern Recognition and Image Analysis, 2017, 27 (3). — p. 387–394.

Об улучшенной оценке меры кластерной структуры в компактном метрическом пространстве

Пушняков Алексей Сергеевич

pushnyakovalex@mail.ru

Москва, Московский физико-технический институт (ГУ)

Рассматривается компактное метрическое пространство (X, ρ) с ограниченной борелевской мерой μ . Под r -кластером понимается любое измеримое множество диаметра не более r . Набор, состоящий из k $2r$ -кластеров, назовем r -кластерной структурой порядка k , если любые два кластера набора отделены на расстояние не менее r . Ранее было показано, что при определенных параметрических ограничениях на распределение расстояний существует r -кластерная структура максимальной меры \mathcal{X}^* , мера которой близка к мере всего пространства. Однако полученная оценка имеет низкую скорость сходимости и неупрощаема в асимптотическом смысле. Предлагается ввести дополнительное ограничение, суть которого состоит в балансировке мер кластеров в искомой структуре. Набор точек (x_1, \dots, x_k) назовем антикликкой порядка k , если $\rho(x_i, x_j) > r$ при всех $1 \leq i < j \leq k$. Мы потребуем, чтобы в нашем метрическом пространстве (X, ρ) мера антиклик была ограничена снизу в следующем смысле:

$$\mu\{(x_1, \dots, x_k) \in X^k : \rho(x_i, x_j) > r, 1 \leq i < j \leq k\} \geq \alpha \mu(X)^k \quad (1)$$

Два предыдущих ограничения состоят в том, что мера пар точек, расстояние между которыми лежит в промежутке $(r, 3r]$, ограничена сверху величиной $\delta \mu(X)^2$ и что мера антиклик порядка $k + 1$ ограничена сверху величиной $\beta \mu(X)^{k+1}$. В купе с неравенством (1) была получена следующая оценка:

$$\frac{\mu(\mathcal{X}^*)}{\mu(X)} \geq 1 - \sqrt{\delta}(2k+1) - \frac{k!(k+2)\beta}{\alpha - \frac{1}{2}k^3\lambda},$$

$$\lambda = \frac{k+1}{2}\delta + \frac{(k+1)^2\beta^2}{2\alpha^2}$$

- [1] Пушняков А. С. Об улучшенной оценке меры кластерной структуры в компактном метрическом пространстве // Журнал вычислительной математики и математической физики, 2017 (на рецензии).

On improved bound for measure of cluster structure in compact metric spaces

Pushnyakov Alexey

pushnyakovalalex@mail.ru

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

A compact metric space (X, ρ) is given. Let μ be a Borel measure on X . By r -cluster we mean a measurable subset of X with diameter at most r . A family of k $2r$ -clusters is called a r -cluster structure of order k if any two clusters from the family are separated by a distance at least r . By measure of a cluster structure we mean a sum of clusters measures from the cluster structure. A sequence (x_1, \dots, x_k) is called an *anticlique of order k* if $\rho(x_i, x_j) > r$ for all $1 \leq i < j \leq k$. In our previous work we showed that under some parametric restrictions for distance distribution measure of maximal cluster structure $\mu(\mathcal{X})^*$ is close $\mu(X)$ and lower bound for $\mu(\mathcal{X})^*$ converges to $\mu(X)$ when corresponding parameters tend to 0. The first condition was that measure of pairs $(x, y) \in X^2$ such that $\rho(x, y) \in (r, 3r]$ is no greater than $\delta\mu(X)^2$. The second condition was that measure of anticliques of order $k + 1$ is no greater than $\beta\mu(X)^{k+1}$. However, this bound asymptotically unimprovable. We propose an additional restriction for distance distribution that is responsible for balance of cluster's measure in cluster structure. Suppose, we have following conditions for measure of anticliques of order k :

$$\mu\{(x_1, \dots, x_k) \in X^k : \rho(x_i, x_j) > r, 1 \leq i < j \leq k\} \geq \alpha\mu(X)^k$$

Under conditions mentioned above we obtained a following bound:

$$\frac{\mu(\mathcal{X}^*)}{\mu(X)} \geq 1 - \sqrt{\delta}(2k + 1) - \frac{k!(k + 2)\beta}{\alpha - \frac{1}{2}k^3\lambda},$$

$$\lambda = \frac{k + 1}{2}\delta + \frac{(k + 1)^2\beta^2}{2\alpha^2}$$

- [1] *Pushnyakov A.* On improved bound for measure of cluster structure in compact metric spaces // Computational Mathematics and Mathematical Physics, 2017 (on review).

О принципе минимизации эмпирического риска на основе усредняющих агрегирующих функций

Шибзухов Заур Мухадинович^{1,2}

szport@gmail.com

¹Москва, Центр «Антистихия» МЧС России

²Нальчик, Кабардино-Балкарский научный центр РАН

В работе предлагается расширение принципа минимизации эмпирического риска на основе применения усредняющих агрегирующих функций (М-средних) для эмпирической оценки средних потерь:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} M_{\rho}\{\ell_1(\mathbf{w}), \dots, \ell_N(\mathbf{w})\},$$

где $\ell_k(\mathbf{w})$ – функция потерь, связанная с k -ым прецедентом.

М-среднее определяется следующим образом:

$$M_{\rho}\{z_1, \dots, z_N\} = \arg \min_u \sum_{k=1}^N \rho(z_k, u),$$

где $\rho(z, u) = g(h(z) - h(u))$, g – выпуклая, h – строго монотонная.

Предложен алгоритм IRLAL (*Iteratively Reweighted Least Averaged Losses*) для поиска \mathbf{w}^* :

$t \leftarrow 0$

Инициализация \mathbf{w}_0

повторять

$$u_t \leftarrow M_{\rho}\{\ell_1(\mathbf{w}_t), \dots, \ell_N(\mathbf{w}_t)\}$$

$$v_k \leftarrow \frac{-\rho''_{uz}(\ell_k(\mathbf{w}_t), u_t)}{\rho''_{uu}(\ell_1(\mathbf{w}_t), u_t) + \dots + \rho''_{uu}(\ell_N(\mathbf{w}_t), u_t)} \quad k = 1, \dots, N$$

$$\mathbf{w}_{t+1} \leftarrow \arg \min_{\mathbf{w}} \sum_{k=1}^N v_k \ell_k(\mathbf{w})$$

$t \leftarrow t + 1$

пока $\{u_t\}$ и $\{\mathbf{w}_t\}$ не стабилизируются

Работа поддержана грантом РФФИ № 15-01-03381 и проектом ОНИТ РАН.

[1] *Шибзухов З. М.* О принципе минимизации эмпирического риска на основе усредняющих агрегирующих функций // Доклады РАН, Москва: МАИК, 2017. Т.476, №5. (принята к публикации)

On the principle of empirical risk minimization based on averaging aggregation functions

Shibzukhov Zaur^{1,2}

szport@gmail.com

¹Moscow, Center “Antistikhiya” Emercom Russia

²Nalchik, Kabardino-Balkarian Scientific Center RAS

In this work we propose an extension of the principle of minimizing empirical risk based on the application of averaging aggregation functions (M-averages) for empirical estimation of average losses:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} M_{\rho}\{\ell_1(\mathbf{w}), \dots, \ell_N(\mathbf{w})\},$$

where $\ell_k(\mathbf{w})$ — losses function for k -th precedent.

M-average is defined as follows:

$$M_{\rho}\{z_1, \dots, z_N\} = \arg \min_u \sum_{k=1}^N \rho(z_k, u),$$

where $\rho(z, u) = g(h(z) - h(u))$, g – convex, h – strictly monotone.

Algorithm IRLAL (*Iteratively Reweighted Least Averaged Losses*) is proposed for searching \mathbf{w}^* :

$t \leftarrow 0$

Initialization of \mathbf{w}_0

repeat

$$u_t \leftarrow M_{\rho}\{\ell_1(\mathbf{w}_t), \dots, \ell_N(\mathbf{w}_t)\}$$

$$v_k \leftarrow \frac{-\rho''_{uz}(\ell_k(\mathbf{w}_t), u_t)}{\rho''_{uu}(\ell_1(\mathbf{w}_t), u_t) + \dots + \rho''_{uu}(\ell_N(\mathbf{w}_t), u_t)}, \quad k = 1, \dots, N$$

$$\mathbf{w}_{t+1} \leftarrow \arg \min_{\mathbf{w}} \sum_{k=1}^N v_k \ell_k(\mathbf{w})$$

$t \leftarrow t + 1$

until $\{u_t\}$ and $\{\mathbf{w}_t\}$ not converge

Work is supported by grant of RFBR 15-01-03381 and project of DNIT RAS.

- [1] Shibzukhov Z. On the Principle of Empirical Risk Minimization Based on Averaging Aggregation Functions // Doklady Mathematics, 2017, Vol.96, No.2, Pleiades Publishing, Ltd., 2017.

Интеллектуальная система поддержки принятия решений с когнитивной компонентой, основанная на отказоустойчивых диагностических тестах

Янковская Анна Ефимовна^{1,2,3,4} ayuankov@gmail.com
Ямшанов Артем Вячеславович^{2*} yav@keva.tusur.ru

¹Томск, Томский Государственный Архитектурно-Строительный Университет

²Томск, Томский государственный университет систем управления и радиоэлектроники

³Томск, Томский государственный университет

⁴Томск, Томский политехнический университет

Предлагается интеллектуальная система поддержки принятия решений (ИС ППР), основанная на интеллектуальном инструментальном средстве ИМСЛОГ (ИИС ИМСЛОГ) с матричным представлением данных и знаний с добавлением ряда существенных модификаций. ИС ППР предназначена для выявления различного рода закономерностей на основе тестовых методов распознавания образов, включая отказоустойчивые диагностические тесты; для поддержки принятия и принятия диагностических и организационно-управленческих решений и их обоснования с использованием развиваемой когнитивной компоненты. Посредством программных модулей реализован основной функционал ИИС ИМСЛОГ: импорт различных входных форматов, пользовательский интерфейс, ряд различных алгоритмов обработки данных, хранение данных, отображение результатов, в том числе с использованием средств когнитивной графики, экспорт результатов и др. Несмотря на удачный выбор архитектуры, позволившей продуктивно развивать ИИС ИМСЛОГ на протяжении длительного времени, современные требования к программным системам (возможность совместной работы, запуск системы на множестве платформ) и современное развитие ЭВМ (замедление роста частот и рост количества ядер, развитие облачных технологий) привели к необходимости пересмотра ряда ранее принятых архитектурных решений, что и приводится в настоящем докладе.

Работа поддержана грантом РФФИ № 16-07-00859а.

- [1] Янковская А. Е., Ямшанов А. В. Интеллектуальная система поддержки принятия решений с когнитивной компонентой, основанная на отказоустойчивых диагностических тестах // Машинное обучение и анализ данных, 2017.

Intelligent system of decision-making support with cognitive component based on the fault-tolerant diagnostic tests

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

ayyankov@gmail.com

Yamshanov Artem^{2*}

yav@keva.tusur.ru

¹Tomsk, Tomsk State University of Architecture and Building

²Tomsk, Tomsk State University of Control Systems and Radioelectronics

³Tomsk, Tomsk State University

⁴Tomsk, Tomsk Polytechnic University

Intelligent system of decision-making support (IS DMS) based on the intelligent instrumental software IMSLOG (IIS IMSLOG) with matrix representation of data and knowledge and implementation of some important modifications is proposed. Mains goals of IS DMS are revealing of different kind of regularities based on test methods of pattern recognition, including fault-tolerant diagnostic tests; making of diagnostic and management decisions and supporting of decision-making process; justifications of made decisions with cognitive component usage. Main functionality of IIS IMSLOG is implemented as a set of the software modules: import of different input formats; user interface; set of different algorithms for data processing; data storage; visualization of results, including cognitive graphic tools; export of results etc. IIS IMSLOG architecture allows productive development of the software for extended periods of time, but modern requirements for software systems (collaboration work, big number of actual platforms) and modern development of computers (slowing of the growth of frequencies and the growth of the number of cores) necessitate to reconsider some of architectural decisions to future development that is described in the report.

This research is funded by RFBR, grant 16-07-00859a.

- [1] *Yankovskaya A., Yamshanov A.* Intelligent system of decision-making support with cognitive component based on the fault-tolerant diagnostic tests // *Journal of Machine Learning and Data Analysis*, 2017.

Приближенный полиномиальный алгоритм для задачи очистки и редактирования данных

<i>Агеев Александр Александрович</i> ¹	ageev@math.nsc.ru
<i>Кельманов Александр Васильевич</i> ^{1,2}	kelm@math.nsc.ru
<i>Пяткин Артем Валерьевич</i> ^{1,2}	artem@math.nsc.ru
<i>Хамидуллин Сергей Асгадулович</i> ^{1*}	kham@math.nsc.ru
<i>Шенмайер Владимир Владимирович</i> ¹	shenmaier@mail.ru

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

²Новосибирск, Новосибирский государственный университет

Анализируются математические аспекты одной из фундаментальных проблем анализа данных — поиска (выбора) в совокупности объектов подмножества похожих объектов. Проблема возникает, в частности, в связи с редактированием данных и их очисткой от посторонних (не похожих) элементов. Рассматривается одна из задач, моделирующих эту проблему, а именно — задача поиска в конечном множестве точек евклидова пространства подмножества максимальной мощности, квадратичный разброс точек которого относительно его неизвестного центроида не превосходит заданную долю от квадратичного разброса точек входного множества относительно его центроида.

Доказано, что задача NP-трудна в сильном смысле. Предложен полиномиальный приближенный алгоритм с оценкой $1/2$ его точности. Представлены результаты численного моделирования, демонстрирующие результативность алгоритма.

- [1] Ageev A. A., Kel'manov A. V., Pyatkin A. V., Khamidullin S. A., Shenmaier V. V. Approximation Polynomial Algorithm for the Data Editing and Data Cleaning Problem // Pattern Recognition and Image Analysis, Vol. 27, № 3. 2017. — P. 365–370.

Approximation Polynomial Algorithm for the Data Editing and Data Cleaning Problem

*Ageev Alexander*¹

ageev@math.nsc.ru

Kel'manov Alexander^{1,2}

kelm@math.nsc.ru

Pyatkin Artem^{1,2}

artem@math.nsc.ru

*Khamidullin Sergey*¹*

kham@math.nsc.ru

*Shenmaier Vladimir*¹

shenmaier@mail.ru

¹Novosibirsk, Sobolev Institute of Mathematics SB RAS

²Novosibirsk, Novosibirsk State University

The work considers the mathematical aspects of one of the most fundamental problems of data analysis: search (choice) among a collection of objects for a subset of similar ones. In particular, the problem appears in connection with data editing and cleaning (removal of irrelevant (not similar) elements). We consider the model of this problem, i.e., the problem of searching for a subset of maximal cardinality in a finite set of points of the Euclidean space for which quadratic variation of points with respect to its unknown centroid does not exceed a given fraction of the quadratic variation of points of the input set with respect to its centroid.

It is proved that the problem is strongly NP-hard. A polynomial $1/2$ -approximation algorithm is proposed. The results of the numerical simulation demonstrating the effectiveness of the algorithm are presented.

This research is funded by RFBR, grants 15-01-00462, 15-01-00976 and 16-07-00168, and by the grant of Presidium RAS (program 5, project 227).

- [1] Ageev A. A., Kel'manov A. V., Pyatkin A. V., Khamidullin S. A., Shenmaier V. V. Approximation Polynomial Algorithm for the Data Editing and Data Cleaning Problem // Pattern Recognition and Image Analysis, Vol. 27, No 3. 2017. — P. 365–370.

Задачи дискретной оптимизации типа многих коммивояжеров

*Германчук Мария Сергеевна*¹

m.german4uk@yandex.ru

*Козлова Маргарита Геннадьевна*¹

art-inf@mail.ru

Лукьяненко Владимир Андреевич^{1*}

art-inf@yandex.ru

¹Симферополь, КФУ им. В. И. Вернадского

Для реальных систем актуальной является проблема анализа и синтеза оптимальных потоков различной природы: ресурсных, информационных и других. В качестве математических моделей используются сети — графовые структуры, вершинам и дугам которых приписаны некоторые величины. Возникает многообразие классов задач дискретной оптимизации (ДО), как правило, NP-полных. Естественный учет информации, связанной с данными задачами ДО, позволяет строить алгоритмы (приближенные, эвристические), пригодные для сложных задач большой размерности. Характерными и тестовыми являются задачи маршрутизации, задачи типа многих коммивояжеров. В работе, используя знаниеориентированный подход, приведен обзор существующих результатов по задачам ДО типа коммивояжера с ограничениями и получены новые постановки задач, предложены алгоритмы их решения. Показано, что учет знаний о структуре сети, целях коммивояжеров, запретах приводит к декомпозиционным (кластерным) алгоритмам. Дальнейшее развитие связывается с подходом, основанным на использовании управляемых интеллектуальных агентов (агентов-коммивояжеров в частности). Рассмотрены обобщенные многоагентные задачи типа коммивояжера, в которых учитываются разнообразные знания, информация, данные необходимые для интеллектуального управления агентами, так и локального управления самого агента, алгоритмы декомпозиции, кластеризации, анализа и синтеза сети. Предварительные численные расчеты подтверждают необходимость создания широкого комплекса алгоритмов, участвующих в оптимальной композиции метаэвристик и наполняющих системы управления.

- [1] *Германчук М. С.* Использование дополнительной информации в задачах дискретной оптимизации типа многих коммивояжеров // Таврический вестник информатики и математики. — 2016. — № 4 (33). — С. 68-82.

The problems of discrete optimization such as Multiple Traveling Salesman Problems

*Germanchuk Mariia*¹

m.german4uk@yandex.ru

*Kozlova Margarita*¹

art-inf@mail.ru

*Lukianenko Vladimir*¹★

art-inf@yandex.ru

¹Simferopol, V. I. Vernadsky Crimean Federal University

The problem of analysis and synthesis of optimal flows like the flow of resources, information flow etc. have a vital importance for the real systems. As a rule, the mathematical models are given by different types of nets. These nets are graph structures with marked vertices and edges. They yield a number of Discrete Optimization (DO) problems most of which are NP-complete. If the information connected with a given DO problems is taken into account, then it is possible to design approximate and heuristic algorithms capable to manipulate complex large-scale data. The Multiple Traveling Salesman Problem (MTSP) and its more generalized version the Vehicle Routing Problem (VRP) are the most typical test problems. In this work on the base of knowledge-oriented approach, we give an overview of existing results for discrete optimization problems such as TSP with restrictions, formulate new problems, and suggest algorithms to solve them. It is shown that the knowledge consideration about the network structure, salesmen objectives, and prohibitions leads to decomposition (cluster) algorithms. The further development is associated with the approach based on the usage of controlled intelligent agents (in particular, salesman agents). We consider the generalized multi-agent traveler salesmen problems taking into account a variety of knowledge, information, data needed for both intelligent agents control and local agent control, the algorithms of decomposition, clustering, analysis and synthesis of networks. The preliminary numerical calculations confirm a necessity in a wide range of algorithms involved in the optimal composition of met heuristics and filling control systems.

- [1] *Germanchuk M.* Information Exploration for Discrete Optimization Problems such as Multiple Traveling Salesman Problems // Taurida Journal of Computer Science Theory and Mathematics. — 2016. — 4 (33). — P. 68-82.

О реализации асимптотически точного подхода к решению некоторых трудных комбинаторных задач

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

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

²Новосибирск, Новосибирский государственный университет

В докладе представлен ряд успешных реализаций асимптотически точного подхода к решению некоторых трудных больше-размерных задач комбинаторной оптимизации, в которых автор за последнее десятилетие принимал непосредственное участие.

1. Новый асимптотически точный алгоритм для Euclidean maximum TSP.

2. m -Peripatetic Salesman Problem (m -PSP).

2.1. Euclidean maximum m -PSP.

2.2. Minimum m -PSP на непрерывных случайных входах.

2.3. Minimum m -PSP на дискретных случайных входах.

2.3 Random maximum m -PSP.

3. Euclidean связный k -фактор максимального веса.

4. Взвешенная задача покрытия графа заданным числом несмежных циклов (m -cycles cover problem).

4.1. Euclidean m -Cycles Cover на максимум.

4.2. UNI(0,1) m -Cycles Cover на минимум.

5. Рандомизированный алгоритм для задачи поиска подмножества векторов максимального веса.

6. Много-индексная задача о назначении на случайных входах.

7. Задача поиска k реберно-непересекающихся остовных деревьев с диаметром, ограниченным снизу.

Исследования задач 1, 2.2 и 2.3 были частично поддержаны РФФ (грант номер 16-11-10041). Исследования остальных задач поддержаны РФФИ (проекты 16-07-00168 и 15-01-00976), а также грантом Президиума РАН (программа 8, проект 227).

[1] *Gimadi E.* Implementation of Asymptotically Optimal Approach to Solving Some Hard Combinatorial Problems // Proc. of the 8th Int. Conf. "Optimization and Applications" (OPTIMA-2017), Petrovac, Montenegro, Oct. 2 – Oct. 6, 2017 (accepted).

On Asymptotically Optimal Approach to Solving Some Hard Combinatorial Problems

Gimadi E.^{1,2}

`gimadi@math.nsc.ru`

¹Novosibirsk, Sobolev Institute of Mathematics SB RAS

²Novosibirsk, Novosibirsk State University

The report presents certain successful examples of the implementation of an asymptotically optimal approach to the solving some large-scale combinatorial problems in which the author has been directly involved in the past half century.

1. A new asymptotically optimal algorithm for the Euclidean maximum TSP.

2. m -Peripatetic Salesman Problem (m -PSP).

2.1. Euclidean maximum m -PSP.

2.2. Random minimum m -PSP with different weight functions.

2.3 Random minimum m -PSP with identical weight functions.

2.3 Random maximum m -PSP.

3. Euclidean maximum connected k -factor.

4. Covering a graph by given number of nonadjacent cycles.

4.1. Euclidean maximum m -Cycles Cover.

4.2. UNI(0,1) minimum m -Cycles Cover.

5. Efficient randomized algorithm for a vector subset problem.

6. Random multi-index Assignment Problem.

7. A minimum k -spanning trees problem with diameter bounded from below.

The study of the problems 1, 2.2 and 2.3 was partially supported by the Russian Science Foundation (project 16-11-10041). The study of another problems was partially supported by the the Russian Foundation for Basic Research (projects 16-07-00168 and 15-01-00976), and the grant of Presidium of RAS (program 8, project 227).

- [1] *Gimadi E.* Implementation of Asymptotically Optimal Approach to Solving Some Hard Combinatorial Problems // Proc. of the 8th Int. Conf. "Optimization and Applications" (OPTIMA-2017), Petrovac, Montenegro, Oct. 2 – Oct. 6, 2017 (accepted).

Алгоритм монотонизации выборки с одновременным отбором объектов и признаков

Зухба Анастасия Викторовна

a__l@mail.ru

Москва, Московский физико-технический институт (ГУ)

Монотонными называют классификаторы, являющиеся монотонными функциями над пространством признаковых описаний объектов. Явный учёт априорных предположений о монотонном характере зависимости позволяет повысить обобщающую способность модели классификации. Большинство методов построения монотонных классификаторов требуют предварительной монотонизации обучающей выборки. Монотонизация сводится к поиску подпространства признаков и подмножества объектов обучающей выборки, монотонных в этом подпространстве.

Если рассматривать задачи отбора объектов и признаков последовательно, возникает следующая проблема: неинформативные признаки мешают отбирать объекты, делая почти все объекты обучающей выборки несравнимыми, а шумовые объекты могут не позволить удалить неинформативные признаки без нарушения условия монотонности.

Предлагаемый алгоритм монотонизации с одновременным отбором объектов и признаков позволяет частично решить эту проблему. На каждой итерации алгоритм выбирает одно из следующих действий: удалить один из объектов, удалить один из признаков, вернуть один из ранее удаленных объектов, вернуть один из ранее удаленных признаков. Выбор осуществляется исходя из жадной оптимизации функционала монотонности выборки. В работе предложены несколько функционалов монотонности выборки, учитывающих ее размер по сравнению с изначальной.

Эксперименты проводились на задаче диагностики заболеваний по электрокардиограмме. Монотонизация выборки производилась предложенным в работе алгоритмом. Качество классификаторов, обученных на полученной подвыборке, сопоставимо с качеством наиболее удачных для данной задачи сочетаний алгоритмов отбора признаков, отбора объектов и классификации.

- [1] *Зухба А. В.* Алгоритм монотонизации выборки с одновременным отбором объектов и признаков // Машинное обучение и анализ данных, 2017.

Monotonization algorithm with simultaneous feature and prototype selection

Zukhba Anastasia

a_l@mail.ru

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

A classifier is isotonic if it is an isotonic function over the space of feature descriptions of the objects. The explicit consideration of a priori assumptions about the isotonic nature of the dependence increases the generalizing ability of the classification model. Most of isotonic classifiers require preliminary monotonization of the training set. Monotonization is prototype and feature selection, aimed at obtaining training subset that satisfies the monotonicity constraint.

If we consider the problems of prototype and feature selection separately, the following obstacle arises: non-informative features complicate the prototype selection, making almost all the objects of the training set incomparable, while noise objects can prevent removal of non-informative features without violating the monotonicity constraint.

We propose the monotonization algorithm with simultaneous feature and prototype selection, which can partially solve the problem. At each iteration, the algorithm chooses one of the following steps: to delete one of the objects, to delete one of the features, to return one of the previously deleted objects, to return one of the previously deleted features. The choice is based on greedy optimization of the monotonicity functional of the training subset. We propose several monotonicity functionals of a training subset, that take into account training subset size.

The experiments was performed on the problem of medical diagnostics using electrocardiogram. We used the algorithm proposed in the paper to perform training set monotonization. The quality of classifiers trained on the resulting subset is comparable to the quality of the most successful combination of feature selection, prototype selection and classification algorithms.

- [1] *Zukhba A.* Monotonization algorithm with simultaneous feature and prototype selection // *Journal of Machine Learning and Data Analysis*, 2017.

О некоторых задачах кластеризации

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

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

²Новосибирск, Новосибирский государственный университет

В докладе дается обзор недавних результатов о сложности и алгоритмах с оценками качества для следующих задач.

Задача 1. *Подмножество точек наибольшей мощности при ограничении на суммарный квадратичный разброс.*

Задача 2. *Подмножество векторов наибольшей мощности при ограничении на нормированную длину суммы векторов.*

Задача 3. *Мощностно-взвешенная квадратичная 2-кластеризация при заданном центре.*

Задача 4. *Взвешенная квадратичная 2-кластеризация при заданном центре.*

Задача 5. *Кластеризация по нормированному критерию k -средних.*

Задача 6. *Кластеризация по нормированному критерию k -средних при заданном центре.*

Задача 7. *Кластеризация по минимуму суммы нормированных квадратов норм.*

Задача 8. *Кластеризация по минимуму суммы квадратов норм.*

Задача 9. *Кластеризация по минимуму суммы норм.*

Задача 10. *Поиск подпоследовательности заданной длины.*

Задача 11. *2-кластеризация последовательности при заданном центре.*

Задача 12. *Кластеризация последовательности при заданном центре одного кластера с ограничениями на мощности.*

Задача 13. *Кластеризация последовательности при заданном центре одного кластера.*

Исследование проблем 1–3, 5, 6, 10, 11, 12 поддержано РФФИ (гранты 15-01-00462, 16-07-00168). Исследование проблем 4, 7, 9, 13 поддержано РФФИ (грант 16-11-10041).

- [1] *Kel'manov A.* On Some Euclidean Partitioning Problems: NP-Hardness and Efficient Approximation Algorithms // Proc. of the 8th Int. Conf. "Optimization and Applications" (OPTIMA-2017), Petrovac, Montenegro, Oct. 2 – Oct. 6, 2017 (accepted).

On some clustering problems

Kel'manov Alexander^{1,2}

kelm@math.nsc.ru

¹Novosibirsk, Sobolev Institute of Mathematics SB RAS

²Novosibirsk, Novosibirsk State University

In the report, we present a short survey on recent results for some clustering problems. Below is a list of considered problems.

Problem 1. *Subset of points with the largest cardinality under constraint on the total quadratic variation.*

Problem 2. *Subset of vectors with the largest cardinality under constraint on normalized length of vectors sum.*

Problem 3. *Cardinality-weighted variance-based 2-clustering with given center.*

Problem 4. *Weighted variance-based 2-clustering with given center.*

Problem 5. *Normalized K-means clustering.*

Problem 6. *Normalized K-Means clustering with a given center.*

Problem 7. *Minimum sum of normalized squares of norms clustering.*

Problem 8. *Minimum sum of squared norms clustering.*

Problem 9. *Minimum sum-of-norms clustering.*

Problem 10. *Finding a subsequence in a sequence.*

Problem 11. *Minimum sum-of-squares 2-clustering problem on sequence with given center of one cluster.*

Problem 12. *Minimum sum-of-squares clustering problem on sequence with given center of one cluster and cluster cardinalities.*

Problem 13. *Minimum sum-of-squares clustering problem on sequence with given center of one cluster.*

The study of problems 1–3, 5, 6, 10, 11 and 12 was supported by the the Russian Foundation for Basic Research (projects 15-01-00462, 16-31-00186, 16-07-00168), and by the Ministry of Science and Education of the Russian Federation under the 5-100 Excellence Programme. The study of problems 4, 7, 9 and 13 was supported by the Russian Science Foundation (project 16-11-10041).

- [1] *Kel'manov A.* On Some Euclidean Partitioning Problems: NP-Hardness and Efficient Approximation Algorithms // Proc. of the 8th Int. Conf. "Optimization and Applications" (OPTIMA-2017), Petrovac, Montenegro, Oct. 2 – Oct. 6, 2017 (accepted).

Декомпозиция разреженных матриц в задачах целочисленного программирования

Лемтюжникова Дарья Владимировна darabbt@gmail.com
Москва, Московский авиационный институт (НИУ)

Разреженные матрицы встречаются в многих приложениях больших данных. Проблематикой таких матриц является хранение, структурирование и декомпозиция соответствующих задач. Одним из подходов является выделение квазиблочной структуры в матрице, которая устроена следующим образом: ненулевые компоненты в матрице объединяются в блоки так, чтобы перемины между блоками не были слишком большими. Это позволяет легко находить декомпозицию исходной задачи и разбивать её на подзадачи согласно блокам. Целью данного исследования является понижение размерности матриц для задач целочисленного программирования, а также изучение порядка решения подзадач на сходимость локального элиминационного алгоритма (ЛЭА).

Сформулирован ряд теорем о выделении квазиблочной структуры в разреженной матрице, а также о связи степени квазиблочной структуры и числа её блоков в зависимости от размерности матрицы и числа ненулевых элементов в ней. Разработана техника понижения размерности разреженных матриц и соответствующих задач целочисленного линейного программирования большой размерности за счёт выделения квазиблочных структур и последующей их обработки. Сформирован метод выделения квазиблочных структур для разреженных матриц, разработаны и реализованы алгоритмы выделения таких структур. Проведены исследования модификаций таких алгоритмов и получены результаты для пяти разных классов матриц. Изучается проблема оптимальной последовательности исключения переменных в ЛЭА. Для этого сформулирован ряд понятий и доказаны свойства графовых структур, соответствующих порядку решения подзадач. Проведено тестирование для различных порядков исключения переменных.

Работа поддержана грантом РФФИ № 16-51-55019.

- [1] *Лемтюжникова Д. В., Ковков Д. В.* Декомпозиция в многомерных задачах с разреженными матрицами // Известия РАН. Теория и Системы Управления, М., 2018, №1.

Decomposition of sparse matrices in integer programming problems

Lemtyuzhnikova Darya

darabbt@gmail.com

Moskow, Moscow Aviation Institute (NRU)

Sparse matrices are found in many large data applications. The problem of such matrices is the storage, structuring and decomposition of the corresponding tasks. One approach is to isolate a quasi-block structure in the matrix, which is organized as follows: non-zero components in matrices are combined into blocks so that the separators between the blocks are not too large. This makes it easy to find the decomposition of the original problem and break it into sub-tasks according to the blocks. The purpose of this study is to reduce the dimension of matrices for integer programming problems, as well as to study the order of solving sub-problems for the convergence of the local elimination algorithm (LEA).

A number of theorems on finding a quasi-block structure in a sparse matrix are formulated, as well as the relation between the degree of a quasiblock structure and the number of its blocks, depending on the dimension of the matrix and the number of nonzero elements in it. A technique for reducing the size of sparse matrices and corresponding problems of integer linear programming of large dimension is developed by separating quasi-block structures and their subsequent processing. A method for isolating quasi-block structures for sparse matrices has been developed, algorithms for isolating such structures have been developed and implemented. Modifications of such algorithms are investigated and results are obtained for five different classes of matrices. The problem of the optimal sequence of elimination of variables in LEA is studied. For this, a number of concepts are formulated and the properties of graph structures corresponding to the order of solving the subtasks are proved. Testing for different orders of exclusion of variables was carried out.

This research is funded by RFBR, grant 16-51-55019.

- [1] *Lemtyuzhnikova D., Kovkov D.* Decomposition in multidimensional problems with sparse matrices // Proceedings of the Russian Academy of Sciences. Theory and Control Systems, M., 2018, 1.

Эффективная разрешимость обобщенной задачи коммивояжера в классах квази- и псевдопирамидальных маршрутов

Хачай Михаил Юрьевич^{1,2}*

mkhachay@imm.uran.ru

Незнахина Екатерина Дмитриевна^{1,2}

eneznakhina@yandex.ru

¹Екатеринбург, Институт математики и механики
им. Н. Н. Красовского УрО РАН

²Екатеринбург, Уральский федеральный университет

Обобщенная задача коммивояжера (Generalized Traveling Salesman Problem, GTSP) является естественным обобщением классической задачи коммивояжера (TSP), постановка которой наряду с полным реберно-взвешенным графом $G = (V, E, w)$ задается разбиением $V = V_1 \cup \dots \cup V_k$ множества его вершин на k кластеров. Цель задачи состоит в поиске циклического маршрута кратчайшего веса, посещающего каждый из кластеров в единственной вершине. Известно, что TSP — NP-трудна в сильном смысле и неаппроксимируема в общей постановке. В то же время, известны многочисленные результаты, посвященные успешному поиску оптимальных (или субоптимальных) решений для различных подклассов задачи. Например, оптимальное решение в классе *пирамидальных* маршрутов может быть найдено за время $O(n^2)$. Введя в рассмотрение *квази-* и *псевдопирамидальные* маршруты для задачи GTSP, мы показали, что оптимальный l -квазипирамидальный и l -псевдопирамидальный маршруты в произвольной постановке задачи GTSP на n вершинах и k кластерах могут быть построены за время $O(4^l n^3)$ и $O(2^l k^{l+4} n^3)$ соответственно. Таким образом показано, что задача GTSP принадлежит классу FPT относительно параметризаций, задаваемых этими типами маршрутов. Кроме того, обоснована полиномиальная разрешимость геометрического подкласса задачи, известного в литературе как GTSP-GC, произвольная постановка которого стеснена дополнительным ограничением $H \leq 2$ на высоту решетки, определяющей кластеры.

Работа поддержана грантом РФФИ № 14-11-00109.

- [1] Хачай М. Ю., Незнахина Е. Д. Разрешимость обобщенной задачи коммивояжера в классе квази- и псевдопирамидальных маршрутов // Труды Института математики и механики УрО РАН, Екатеринбург: ИММ УрО РАН, 2017. Т. 23, no. 3. — С. 280–291.

Efficient solvability of the Quasi- and Pseudo-Pyramidal Generalized Traveling Salesman Problem

Khachay Michael^{1,2*}

mkhachay@imm.uran.ru

Neznakhina Katherine^{1,2}

eneznakhina@yandex.ru

¹Ekaterinburg, Krasovskii Institute of Mathematics and Mechanics
UB RAS

²Ekaterinburg, Ural Federal University

The Generalized Traveling Salesman Problem (GTSP) is a natural extension of the famous Traveling Salesman Problem (TSP), whose instance is defined by complete edge-weighted graph $G = (V, E, w)$ and a partition $V = V_1 \cup \dots \cup V_k$ of its nodeset to k clusters. The goal is to find a minimum weight cyclic tour visiting each cluster once.

It is widely known that TSP is strongly NP-hard and inapproximable problem in its general setting. Same time, for more narrow subclasses of TSP, there are exist numerous results on efficient finding optimal (or close suboptimal) solutions. For instance, for any weighting function, the cheapest *pyramidal* tour can be found in time $O(n^2)$.

For GTSP, we introduce *quasi-* and *pseudo-*pyramidal routes and show that l -quasi-pyramidal and l -pseudo-pyramidal routes can be obtained in time of $O(4^l n^3)$ and $O(2^l k^{l+4} n^3)$, respectively.

Thus, we show that, in the most common setting, GTSP belongs to FPT class with respect to parametrizations induced by these routes. Further, as a consequence, we prove polynomial time solvability for one geometric subclass of the problem (in literature it is known as GTSP in Grid Clusters or GTSP-GC), whose clusters are defined by the unit grid in the Euclidean plane with height $H \leq 2$.

Research was supported by RSF, grant 14-11-00109.

- [1] *Khachai M. Yu., Neznakhina E. D.* Solvability of the Generalized Traveling Salesman Problem in the class of quasi- and pseudo-pyramidal tours // Proc. of the Inst. of Math. and Mech. UrB RAS, 2017. Vol. 23, no. 3. — p. 280–291 (in Russian).

Гарантированная оценка качества кластеризации методом k -медиан на отрезке

Хачай Михаил Юрьевич^{1,2*}

mkhachay@imm.uran.ru

Панкратов Василий Сергеевич¹

pankratov.vs@gmail.com

¹Екатеринбург, Институт математики и механики
им. Н. Н. Красовского УрО РАН

²Екатеринбург, Уральский федеральный университет

Задача о k -медианах является одной из наиболее известных конкретизаций метрической задачи кластеризации, чья постановка может быть описана следующим образом. Для заданного числа $k \geq 1$ и конечной выборки $\xi = (x_1, \dots, x_n)$ из метрического пространства (X, ρ) требуется найти разбиение $\{1, \dots, n\}$ на k кластеров C_1, \dots, C_k , указав для j -го кластера центр c_j так, что

$$\sum_{j=1}^k \sum_{i \in C_j} \rho(x_i, c_j) = \sum_{i=1}^n \min\{\rho(x_i, c_1), \dots, \rho(x_i, c_k)\} \rightarrow \min. \quad (1)$$

При произвольном заданном разбиении центры кластеров определяются однозначно по формуле $c_j = \arg \min\{\sum_{i \in C_j} \rho(x_i, c) : c \in X\}$, т.е. c_j совпадает с медианой подвыборки $\xi_j = (x_i : i \in C_j)$.

Известно, что задача о k -медианах труднорешаема даже для евклидовой метрики и не имеет в общем случае РТАС (при условии $P \neq NP$). Тем не менее, в \mathbb{R}^d при каждом фиксированном k задача обладает рандомизированной приближенной схемой с трудоемкостью $O(2^{(k/\varepsilon)^{O(1)}} \cdot dn)$, а на вещественной прямой задача полиномиально разрешима.

Наряду с эффективными алгоритмами важным представляется обоснование гарантированных оценок точности наилучшей кластеризации, зависящих исключительно от длины выборки и числа кластеров. В данной работе мы показываем, что произвольная выборка длины n из отрезка $[0, 1]$ может быть разбита на k кластеров так, что значение функционала (1) не превысит $0.5n/(2k - 1)$, причем данная оценка является достижимой.

Работа поддержана грантом РФФИ № 16-07-00266.

- [1] Khachay M., Pankratov V., and Khachay D. Attainable best guarantee for the accuracy of k -medians clustering in $[0, 1]$ // CEUR Workshop Proc., Aachen: RWTH, 2017.

Attainable best accuracy guarantee for k -medians clustering in $[0, 1]$

Khachay Michael^{1,2,*}

mkhachay@imm.uran.ru

*Pankratov Vasiliy*¹

pankratov.vs@gmail.com

¹Ekaterinburg, Krasovskii Institute of Mathematics and Mechanics
UB RAS

²Ekaterinburg, Ural Federal University

The k -medians problem is a famous setting of center-based clustering problems, whose instance can be defined as follows. For a given number $k \geq 1$ and a given sample $\xi = (x_1, \dots, x_n)$ taken from metric space (X, ρ) it is required to find a partition $\{1, \dots, n\}$ onto k clusters C_1, \dots, C_k and, for any j -th cluster, to point out a center c_j such that

$$\sum_{j=1}^k \sum_{i \in C_j} \rho(x_i, c_j) = \sum_{i=1}^n \min\{\rho(x_i, c_1), \dots, \rho(x_i, c_k)\} \rightarrow \min. \quad (1)$$

For an arbitrary partition, cluster centers are defined by the formula $c_j = \arg \min\{\sum_{i \in C_j} \rho(x_i, c) : c \in X\}$, i.e. c_j coincides with *the median* of the subsample $\xi_j = (x_i : i \in C_j)$.

The k -medians clustering problem is NP-hard even for the Euclidean metric and has no PTAS (unless $P = NP$). Nevertheless, in \mathbb{R}^d , for any fixed k , the problem has a randomized approximation scheme with time complexity $O(2^{(k/\varepsilon)^{O(1)}} \cdot dn)$. Furthermore, the problem is polynomially solvable in the real line. Besides the design of efficient algorithms, it seems valuable to prove theoretical accuracy bounds depending ultimately on sample length and the number of clusters.

In this paper, we show that an arbitrary n -length sample in $[0, 1]$ can be partitioned into k clusters such that the value of functional (1) does not exceed $0.5n/(2k - 1)$. This bound is attainable.

Research was supported by RFBR, grant 16-07-00266.

- [1] *Khachay M., Pankratov V., and Khachay D.* Attainable best guarantee for the accuracy of k -medians clustering in $[0, 1]$ // CEUR Workshop Proc., Aachen: RWTH, 2017.

Отбор подмножества активных биржевых индексов в динамической модели инвестиционного портфеля

Бабичев Дмитрий Олегович^{1*} shiftdmitryalt@gmail.com
*Красоткина Ольга Вячеславовна*¹ o.v.krasotkina@yandex.ru
*Моттль Вадим Вячеславович*² vmottl@yandex.ru

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

²Москва, Федеральный исследовательский центр «Информатика и управление» РАН

Задача ретроспективного анализа скрытой динамики состава инвестиционного портфеля (Dynamic Style Analysis) заключается в оценивании дневной, месячной или квартальной последовательности долевых распределений объема капитала между большим числом секторов экономики, которое рассматривается инвестиционной компанией как ее глубокая тайна. Массивом исходных данных является временной ряд периодических доходностей компании (returns), о которых она обязана отчитываться, вместе с известными рядами биржевых индексов – усредненных доходностей секторов экономики, в которые предположительно может быть вложен капитал (Returns Based Analysis). Такая задача сформулирована как задача оценивания нестационарной регрессии. Число индексов (регрессоров) может составлять сотни и даже тысячи, образуя большой массив данных, в то время как портфель типичной инвестиционной компании, как правило, содержит ценные бумаги из не более, чем десятка секторов экономики. В докладе излагается вероятностная модель динамики долевого состава инвестиционного портфеля, в основе которой лежит априорное нормальное-гамма распределение скрытого марковского случайного процесса. Алгоритм отбора с регулируемой селективностью, имеющий линейную вычислительную сложность по общему числу индексов, находит среди них активное подмножество вместе со степенями волатильности долевых коэффициента вплоть до полной неизменности доли индекса в портфеле.

Работа поддержана грантом РФФИ № 17-07-00993.

- [1] *Krasotkina O., Mottl V., Markow M., et al.* Methods of hyperparameter estimation in time-varying regression models with application to dynamic style analysis of investment portfolios // Lecture Notes in Computer Science, Vol. 10358, Springer, 2017, pp. 431-450.

Selection of the subset of active stock market indices in the dynamic model of an investment portfolio

*Babichev Dmitry*¹★

shiftdmitryalt@gmail.com

*Krasotkina Olga*¹

o.v.krasotkina@yandex.ru

*Mottl Vadim*²

vmottl@yandex.ru

¹Moscow, Lomonosov Moscow State University

²Moscow, Federal Research Center “Computer Science and Control” of RAS

Retrospective analysis of the hidden dynamics of an investment portfolio (Dynamic Style Analysis) implies estimation of daily, monthly or quarterly sequence of capital shares over a large number of economic sectors, which is a secret of the investment company. The data set to be processed is made by the time series of periodic returns on the portfolio, which it is obliged to report daily, along with those on known stock market indices – averaged returns of economic sectors, in which the capital may be supposedly invested (Returns Based Analysis). This problem is formulated as that of non-stationary regression estimation. The number of regressors (stock market indices) may run into hundreds and even thousands, what turns the array of time series into a big data set, whereas the portfolio of a typical investment company holds assets from, as a rule, not more than dozens of economic sectors. In this paper, we propose a probabilistic model of dynamic structure of an investment portfolio. The model is underlied by an a priori gamma-normal distribution of the hidden random process. The respective regressor selection algorithm of supervised selectivity, that has linear computational complexity with respect to the total number of indices, finds among them the most likely active subset and estimates the volatility level for the weighting factor at each selected index, up to the full invariability of its share in the portfolio composition.

This research is funded by RFBR, grant 17-07-00993.

- [1] *Krasotkina O., Mottl V., Markow M., at al.* Methods of hyperparameter estimation in time-varying regression models with application to dynamic style analysis of investment portfolios // Lecture Notes in Computer Science, Vol. 10358, Springer, 2017, pp. 431-450.

Оптимизация сложности моделей анализа временных рядов в экономике

Кириллюк Игорь Леонидович^{1*}

igokir@crambler.ru

*Кузнецова Анна Викторовна*²

azfor@narod.ru

*Сенько Олег Валентинович*³

senkoov@mail.ru

¹Москва, Институт экономики РАН

²Москва, Институт биохимической физики им. Н. М. Эмануэля РАН

³Москва, Федеральный исследовательский центр «Информатика и управление» РАН

Важную роль в экономических исследованиях играет анализ многомерных временных рядов, которые называются панельными данными (ПД). В [1] с использованием метода оптимальных достоверных разбиений, а также с помощью специально разработанных способов учёта эффектов множественного тестирования были выявлены существенные различия между группами стран с различным институциональным типом. Как было показано позднее, важную роль могут играть различия по параметрам производственных функций, характеризующих зависимость выпуска продукции от затрачиваемых факторов производства: труда и капитала. Наряду с широко известной исходной и сводящейся к одномерной регрессии моделью Кобба-Дугласа, существует большое количество альтернативных моделей производственных функций, имеющих, как правило, более высокую сложность, и более высокий уровень параметризации. Естественно, что более сложные модели позволяют лучше аппроксимировать, что в условиях ограниченного периода наблюдений приводит к необходимости учёта эффекта переобучения. Иными словами, возникает вопрос о выборе оптимального уровня сложности. В настоящее время существует целый ряд подходов к оценке, включая критерий Акаике, Байесовский информационный критерий и др. В настоящей работе рассматривается подход, основанный на использовании перестановочных тестов, как позволяющий оценить использование дополнительной параметризации в терминах p -значений.

Работа поддержана грантом РФФИ 17-02-00207.

- [1] *Kirilyuk I. L., Kuznetsova A. V., Sen'ko O. V., Morozov A. M.* Method for detecting significant patterns in panel data analysis. // Pattern recognition and image analysis (Advances in mathematical theory and applications), Springer, 2017. V 27, № 1 — P.94–104.

Optimization of models complexity for time series analysis in economics

*Kirilyuk Igor*¹*

igokir@rambler.ru

*Kuznetsova Anna*²

azfor@narod.ru

*Senko Oleg*³

senkoov@mail.ru

¹Moscow, Institute of Economics, RAS

²Moscow, Emanuel Institute of Biochemical Physics of RAS

³Moscow, Federal Research Center “Computer Science and Control” of RAS

Multidimensional time series or panel data take important place in economic researches. Method of optimal valid partitioning together with special way of multiple testing effect correcting allow to discover significant discrepancies between groups of countries of different institutional types. As it was shown later deviations by production functions parameters between these group are also statistically significant. Production functions relates physical output of a production process on factors of production: capital and labour. Well known Cobb-Douglas function may be reduced to simple regression. But variety of alternative models with greater complexity and parametrization levels exists besides. It is naturally that more complicated models allow to achieve better approximation of data in case of limited periods due to overfitting. So problem of optimal complexity level evaluating arises. Today this problem may be solved with the help of several methods: Akaike criterion, BIC and others. We discuss method based on permutation tests evaluating necessity of additional parametrization in terms of p -values.

Research was supported by RFBR, grant 17-02-00207.

- [1] *Kirilyuk I. L., Kuznetsova A. V., Sen'ko O. V., Morozov A. M.* Method for detecting significant patterns in panel data analysis. // Pattern recognition and image analysis (Advances in mathematical theory and applications), Springer, 2017. V 27, 1 — P. 94–104.

Вероятностное прогнозирование мощности ветряных электростанций с использованием адгоритма K ближайших соседей

Мангалова Екатерина Сергеевна¹ e.s.mangalova@hotmail.com

¹Красноярск, ООО «Ар Ди Сайнс»

Эффективная работа ветряных электростанций предполагает оптимизацию режимов их работы в единой энергетической системе. Прогнозы выработанной мощности каждого ветрогенератора имеют решающее значение при управлении энергетической системой. Вероятностные прогнозы в энергетике являются более предпочтительными, чем точечные прогнозы. Цель вероятностного прогнозирования в ветроэнергетике заключается в предсказании вероятностного распределения (его квантилей) выработанной мощности, нормированной на максимальную мощность станции. Объясняющие переменные — прогнозы ветра в виде зональной и меридиональной компонент.

Алгоритм k ближайших соседей показывает высокую точность при точечных прогнозах мощностей ветряных электростанций, и имеет следующие достоинства: результаты алгоритма k ближайших соседей могут быть интерпретированы экспертом; значимость входных переменных может быть оценена естественным образом на основании весов в мере расстояния; не требуется многократное обучение модели при поступлении новых данных.

Алгоритм k ближайших соседей был адаптирован для решения задачи вероятностного прогнозирования. Предложенный алгоритм состоит из двух последовательных шагов: поиск ближайших соседей, оценка квантилей распределения.

Алгоритм позволяет оценивать вероятностное распределение непосредственно по исходным данным без каких-либо дополнительных предположений, например, о законе распределения прогнозируемой величины.

Предложенный алгоритм занял третье место в Global Energy Forecasting Competition 2014.

- [1] *Mangalova E., Shesterneva O.* K-nearest neighbors for GEFCom2014 probabilistic wind power forecasting // International Journal of Forecasting, Elsevier, 2016. — P.1067–1073. <http://www.sciencedirect.com/science/article/pii/S0169207015001429>.

Probabilistic wind power forecasting using K-nearest neighbors algorithm

*Mangalova Ekaterina*¹

e.s.mangalova@hotmail.com

¹Krasnoyarsk, LLC “RD Science”

The effective operation of wind power plants involves the optimization of their operating modes within the integrated energy system. Predictions of the power production of each individual wind farm are crucial for energy management. For this reason, a probabilistic forecasts are preferable to single-valued forecasts. The aim of the probabilistic wind power forecasting is to forecast the probabilistic distribution (in quantiles) of the wind power generation of wind farms. The target variable is power generation, normalized by the respective nominal capacity. The explanatory variables are the past power measurements and input weather forecasts, given as zonal and meridional components.

The k -nearest neighbors algorithm is highly efficient for spot wind power forecasting, and has several advantages:

- the k -nearest neighbors algorithm allows its results to be interpreted by experts,
- the significance of the input variables can be estimated, provided that a weighted distance metric is used,
- no multiple learning is needed when new data are obtained.

The k -nearest neighbors algorithm is adapted for probabilistic forecasting. The proposed algorithm consists of two steps:

- a nearest neighbors search,
- the estimation of the quantiles.

This algorithm allows estimating the probabilistic distribution from the available data directly, without any assumptions as to the shape of the predictive distribution.

Based on the competition criteria, the predictive model achieved third place in the Global Energy Forecasting Competition 2014.

- [1] *Mangalova E., Shesterneva O.* K-nearest neighbors for GEFCom2014 probabilistic wind power forecasting // International Journal of Forecasting, Elsevier, 2016. — P. 1067–1073. <http://www.sciencedirect.com/science/article/pii/S0169207015001429>.

Алгоритмы анализа динамики стиля инвестиций по большому числу биржевых индексов

Пугач Илья Александрович^{1*}

iliapugach@gmail.com

*Моттль Вадим Вячеславович*¹

vmottl@yandex.ru

*Красоткина Ольга Вячеславовна*²

o.v.krasotkina@yandex.ru

¹Москва, Московский физико-технический институт (ГУ)

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

Динамика дневного, месячного или квартального долевого распределения капитала, вложенного в портфель инвестиционного фонда, принципиально интересна инвесторам, являясь, в то же время, профессиональной тайной фонда. Для наблюдения доступен лишь временной ряд периодических доходностей фонда (returns), о которых он обязан отчитываться. Задача построения ретроспективной модели динамики распределения капитала (стиля инвестиций) на основе анализа этого временного ряда совместно с известными рядами биржевых индексов, т.е. усредненных доходностей секторов экономики, в которые может быть вложен капитал, получила в мировой литературе название Return Based Dynamic Style Analysis (RBDSA). Эта задача решается нами как задача оценивания линейной нестационарной регрессионной модели временного ряда доходностей анализируемого портфеля, в которой в роли множества регрессоров выступает исходная совокупность временных рядов биржевых индексов, число которых может исчисляться сотнями и даже тысячами, образуя большой массив данных. Наша процедура RBDSA имеет линейную вычислительную сложность по числу биржевых индексов, в то время как традиционный подход к задаче регрессионного анализа приводит к кубической вычислительной сложности по этой характеристике массива. Процедура допускает ограничение на неотрицательность динамических коэффициентов регрессии, что отражает специфику анализа паевых (пенсионных) фондов, в которых запрещено хеджирование, т.е. заимствование внешнего капитала.

Работа поддержана грантом РФФИ № 17-07-00993.

- [1] *Krasotkina O., Mottl V., Markow M., et al.* Methods of hyperparameter estimation in time-varying regression models with application to dynamic style analysis of investment portfolios // *Lecture Notes in Computer Science*, Vol. 10358, Springer, 2017, pp. 431-450.

Algorithms for dynamic style analysis of investments from a large number of stock market indices

*Pugach Ilya*¹★

iliapugach@gmail.com

*Mottl Vadim*¹

vmottl@yandex.ru

*Krasotkina Olga*²

o.v.krasotkina@yandex.ru

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

²Moscow, Lomonosov Moscow State University

The dynamics of daily, monthly or quarterly sequence of capital shares over a large number of economic sectors, in which the capital of an investment fund can be placed, is extremely interesting to investors, being, at the same time, a secret of the investment company. What is available for observation, is only the time series of the fund's periodic returns, which it is obliged to report. The problem of building a retrospective dynamic model of capital sharing (Style Analysis) as result of processing this time series along with those of known returns on the stock market indices, i.e., averaged returns of economic sectors, in which the capital may supposedly be invested, is named in the literature as Return Based Dynamic Style Analysis (RBDSA). We solve this problem as that of estimating the linear non-stationary regression model of the time series of periodic returns on the portfolio under analysis. In this model, the set of regressors is the entire collection of stock market indices, which can count up to hundreds and even thousands of indices, and make, thus, a big data set. Our RBDSA procedure has linear computational complexity with respect to the number of stock market indices, whereas the traditional approach to regression analysis results in cubic computational complexity. The procedure allows for the restriction on non-negativity for dynamic regression coefficients. The latter possibility correspond with the specificity of the analysis of mutual (pension) funds, for which hedging, i.e. use of borrowed capital, is prohibited.

This research is funded by RFBR, grant 17-07-00993.

- [1] *Krasotkina O., Mottl V., Markow M., et al.* Methods of hyperparameter estimation in time-varying regression models with application to dynamic style analysis of investment portfolios // Lecture Notes in Computer Science, Vol. 10358, Springer, 2017, pp. 431-450.

Выбор оптимальной модели прогнозирования объемов грузовых железнодорожных перевозок

Рудаков Константин Владимирович^{1*} rudakov@frccsc.ru

*Кузнецов Михаил Павлович*² mikhail.kuznecov@phystech.edu

*Мотренко Анастасия Петровна*¹ anastasiya.motrenko@phystech.edu

*Громов Андрей Николаевич*¹ agromov@forecsys.ru

*Попова Екатерина Витальевна*³ popova_ev2013@mail.ru

Стрижов Вадим Викторович^{1,2} strijov@phystech.edu

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

²Москва, Московский физико-технический институт (ГУ)

³Москва, Институт экономики и развития транспорта

Решается задача выбора оптимальной модели краткосрочного прогнозирования объемов железнодорожных перевозок по историческим и экзогенным временным рядам. Исторические данные содержат информацию об объемах перевозок различных типов грузов между парами станций. Предполагается, что результат выбора оптимальной модели зависит от уровня агрегирования по типам грузов, пунктам отправления и назначения и по времени. Рассмотрены модели векторной авторегрессии, интегрированная модель авторегрессионного скользящего среднего и непараметрическая модель гистограммного прогнозирования. Предложены критерии сравнения прогнозов на основе расстояний между ошибками прогнозов моделей. Данные критерии используются для анализа моделей с целью определения допустимых запросов на прогноз, в том числе фактической глубины прогнозирования.

Исследование поддержано грантами РФФИ № 17-20-01212 и 17-20-01184.

- [1] *Рудаков К.В., Стрижов В.В., Каширин Д.О., Кузнецов М.П., Мотренко А.П., Стенина М.М.* Выбор оптимальной модели прогнозирования объемов грузовых железнодорожных перевозок // Автоматика и телемеханика. 2017, 1. С. 91-105.

On selecting the optimal forecasting model for railway freight transportation demand

Rudakov Konstantin^{1*} rudakov@frccsc.ru
*Kuznetsov Mikhail*² mikhail.kuznecov@phystech.edu
*Motrenko Anastasia*¹ anastasiya.motrenko@phystech.edu
*Gromov Andrey*¹ agromov@forecsys.ru
*Popova Ekaterina*³ popova_ev2013@mail.ru
Strijov Vadim^{1,2} strijov@phystech.edu

¹Moscow, Federal Research Center “Computer Science and Control” of RAS

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

³Moscow, Institute of Economics and Transport Development

We address the problem of optimal model selection for short-term rail freight transportation forecast. The forecast is based on historical time series as well as on exogenous time series used a source of additional information. Historical data contain information about transportation volumes for various freight types between pairs of stations. We argue that the result of model selection depends on level of data aggregation by freight types, destination and origin stations, and on the choice of time scale. We consider vector autoregression model, integrated autoregressive moving average model and a nonparametric model of histogram forecasting. We propose several criteria for model comparison, based on the distance between errors made by forecasting models. The purpose of these criteria is to provide platform for comprehensive analysis of forecasting models. This includes validation of forecasting parameters and time-frame.

This work was supported by the Russian Foundation for Basic Research, awards 17-20-01212 and 17-20-01184.

- [1] *Rudakov K. V., Strizhov V. V., Kashirin D. O., Kuznetsov M. P., Motrenko A. P., Stenina M. M.* Selecting an Optimal Model for Forecasting the Volumes of Railway Goods Transportation // Automation and Remote Control, 2017, Vol. 78, No. 1, 74-87.

Об идентификации статистической модели транспортных потоков с использованием групп АТС

Старожилец Всеволод Михайлович^{1,2*} starvsevol@gmail.com

Чехович Юрий Викторович^{1,2} chehovich@forecsys.ru

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

²Москва, Московский физико-технический институт (ГУ)

В работе предлагается статистическая модель транспортных потоков для моделирования движения транспортных средств на автомагистралях идентифицируемая на данных из гетерогенных источников. Модель симулирует движение групп транспортных средств по магистрали с использованием фундаментальной диаграммы на выбранном участке автодороги для расчета скорости группы. Критерием качества выбрана среднеквадратичная ошибка между предсказанным числом проехавших автотранспортных средств и их реальным числом.

Проводятся три вычислительных эксперимента для подтверждения работоспособности модели, а также анализа её поведения при моделировании ситуации перекрытия одной из полос автомагистрали. В первом и втором эксперименте рассматривается участок МКАД без въездов и съездов в начале и в конце которого установлены дорожные датчики. В первом эксперименте с помощью модели и информации о числе въехавших на участок автомобилей определяется число съехавших транспортных средств и сравнивается с значением, зафиксированным дорожными датчиками. Во втором эксперименте проверяется адекватность поведения модели в ситуации перекрытия одной из полос на съезде с участка автодороги, однако реальных данных для такого случая у нас нет. Третий эксперимент проводится на данных видеосъемки на участке автомагистрали в начале и в конце которого установлены светофоры.

- [1] *Старожилец В. М., Чехович Ю. В.* Об идентификации статистической модели транспортных потоков с использованием групп АТС // Машинное обучение и анализ данных, 2017.

About identification of a statistical model of traffic flows using vehicle groups

Starozhilets Vsevolod^{1,2,*}

starvsevol@gmail.com

Chehovich Yuriy^{1,2}

chehovich@forecsys.ru

¹Moscow, Dorodnicyn Computing Centre of FRC CSC RAS

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

In this paper we propose a statistical model of traffic flows for modeling speed and number of cars on highways identified on data from heterogeneous sources. The model simulates movement of car groups along the highway using fundamental diagrams for road segments to calculate the car groups speed. The criterion of quality is the root-mean-square error between the predicted number of passed vehicles and the actual number of vehicles.

We provide computational experiments to confirm the adequacy of the model, as well as analyzing its behavior in situation of blocking one of the lanes of the highway. In the first and second experiments we consider Moscow Circular Automobile Road segment without entrances and exits. Traffic detectors are located at the beginning and at the end of segment. In the first experiment, using the model and information on the number of vehicles entering the segment, the number of vehicles that have moved out is determined and compared with the value recorded by the traffic detectors. In the second experiment, the adequacy of the model's behavior in the situation of blocking one of the lanes at the end of road segment is checked, but we do not have real data for this case. The third experiment is carried out on the video data on the section of the motorway, where traffic lights located at the beginning and at the end of segment.

- [1] *Starozhilets V., Chehovich Yu.* About identification of a statistical model of traffic flows using vehicle groups // *Journal of Machine Learning and Data Analysis*, 2017.

Поиск плавно меняющихся закономерностей на рынке недвижимости

Филипенков Николай Владимирович^{1*} n.filipenkov@mail.ru
*Петрова Марина Алексеевна*² marina_petrova@mail.ru

¹Москва, САС Институт

²Москва, Национальный исследовательский ядерный университет «МИФИ»

Стартующая в Москве программа реновации жилья делает крайне актуальной задачу определения инвестиционной привлекательности проектов застройки в различных районах Москвы. Одним из ключевых компонентов этой задачи является определение стоимости квартиры по ряду признаков. При этом признаки могут как описывать конкретный объект (например, площадь квартиры, этаж, год постройки и т.д.), так и состояние экономики (динамика ВВП, уровень безработицы, курсы валют и т.д.) на момент сделки по продаже жилья.

Наличие зависимости от времени позволяет применять для данной задачи алгоритм поиска плавно меняющихся закономерностей, разрабатываемый авторами. В работе вводится понятие закономерности и меры сходства закономерностей, выявляются закономерности плавно меняющиеся с течением времени.

Эксперименты в работе проводятся на реальных данных о стоимости жилья в Москве в период с 2011 по 2015 годы, предоставленных. В ходе экспериментов на реальных данных выявлены как универсальные закономерности, не зависящие от состояния экономики, так и закономерности, на которые макроэкономические признаки оказывают существенное влияние.

Работа поддержана грантом РФФИ № 16-07-01156.

- [1] *Филипенков Н. В.* Поиск плавно меняющихся закономерностей на рынке недвижимости // Машинное обучение и анализ данных, 2017.

Mining Slightly Changing Patterns on the Housing Market

*Filipenkov Nikolay*¹★

n.filipenkov@mail.ru

*Petrova Marina*²

marina_petrova@mail.ru

¹Moscow, SAS Institute

²Moscow, National Research Nuclear University MEPhI

House renovation program is starting in Moscow. This makes extremely important the estimation of the profitability of construction projects in the various Moscow districts. One of the key components of this task is the property price estimation by various features. The features may describe the property (square meters, floor, date of construction etc.) or the state of economy at the moment of the property purchase (GDP, unemployment rate, currency exchange rates etc.)

Since there is a time dependency one could apply for this task the slightly changing patterns mining algorithm, which is being developed by the authors. In this paper the pattern and the pattern similarity measure are introduced. The slightly changing patterns are mined.

The experiments were performed on the real property price data in Moscow during the period from 2011 to 2015. During the experiments on the real data there the universal patterns were mined that are not dependent on the economy state. There were also mined the patterns that are significantly dependent on the macroeconomic conditions.

This research is funded by RFBR, grant 16-07-01156.

- [1] *Filipenkov N.* Mining Slightly Changing Patterns on the Housing Market // *Journal of Machine Learning and Data Analysis*, 2017.

Использование методов обработки космических изображений для оценки состояния рекреационной зоны Таганрогского залива

Архипова Ольга Евгеньевна^{1,2*}

arkhipova@sfedu.ru

*Герасюк Виктория Сергеевна*³

gerasyuk.v@gmail.com

¹Ростов-на-Дону, Южный федеральный университет

²Ростов-на-Дону, Институт аридных зон ЮНЦ РАН

³Ростов-на-Дону, Южный научный центр РАН

В ходе выполнения работы был проведен пространственный анализ одновременных мультиспектральных космических снимков береговой зоны Таганрогского залива, период исследования 2006-2016 гг. Для обработки мультиспектральных данных применялся метод классификации [1] с обучением и метод одноканального обнаружения изменений. Метод интерактивной классификации с обучением предполагает, что статистические данные для каждого класса в каждом канале нормально распределены, вычисляется вероятность, что каждый пиксель принадлежит определенному классу. Выбор класса осуществлялся вручную. Для определения эталонных участков, сравнительно небольшой одинаковый участок местности на всех снимках сопоставляется с аналогичным ему фрагментом на топографической карте. Стандартные алгоритмы одноканального обнаружения изменений заключались в вычитании яркостей одного снимка из яркостей другого [1], либо делении яркостей одного снимка на яркости другого. При использовании метода главных компонент оценивалась «типичность» изменения яркостей, а в качестве меры отклонения от «типичности» используется вторая главная компонента, ортогональная первой. Постклассификационная обработка и обнаружение изменений путем составления статистики и расчета карты различий показали, что отмечаются изменения площадей всех выделяемых классов объектов, что связано как с активной хозяйственной деятельностью, так и с особенностью геоморфологии района.

- [1] *Архипова О. Е. Качалина Н. А.* Оценка засоренности антропогенных фитоценозов на основе данных дистанционного зондирования земли (на примере Амброзии полыннолистной) Исследование Земли из космоса, Москва: Издательство Наука, 2014. 6 С. 15–34.

Space-imaging techniques to assess the status of a recreational zone of the Taganrog Bay

Arkhipova Olga^{1,2*}

arkhipova@sfnedu.ru

*Gerasyuk Victoriya*³

gerusyuk.v@gmail.com

¹Rostov-on-Don, Southern Federal University

²Rostov-on-Don, Institute of Arid zones SSC RAS

³Rostov-on-Don, Southern Scientific Center of RAS

In the course of the work, spatial analysis of multispectral Landsat data of the coastal zone of the Taganrog Bay was carried out, the study period is 2006-2016. For the processing of space images, the method of classification with training and the method of single-channel detection of changes were applied. The Interactive Supervised Classification method assumes that the statistics for each class in each channel are normally distributed, the probability that each pixel belongs to a particular class is calculated. For the definition of reference areas, a comparatively small identical area of the terrain in all images is compared with a similar fragment on the topographic map. Standard algorithms for single-channel detection of changes consisted in subtracting the brightness of one image from the brightness of another, or dividing the brightness of one picture by the brightness of the other. When using the principal component method, the "typical" brightness variation was evaluated, and as a measure of deviation from "typicality" the second main component orthogonal to the first one is used. Post classification processing and detection of changes by compiling statistics and calculating a map of differences showed that changes in the areas of all allocated classes of objects are recorded, related both to economic activity and the feature of the geomorphology of the area.

- [1] *Arkhipova O. E. Kachalina N. A.* Assessment of the weediness of anthropogenic phytocenoses based on remote sensing data of the earth (on the example of *Ambrósia artemisiifolia*) *The study of the Earth from space*, Moscow: Nauka, 2014. 6 p. 15–34.

Представление больших данных в задачах поддержки принятия решений по обеспечению устойчивого развития региона

Архипова Ольга Евгеньевна^{1,2}

arkhipova@sfedu.ru

*Сурков Федор Алексеевич*²*

fasurkov@sfedu.ru

¹Ростов-на-Дону, Институт аридных зон ЮНЦ РАН

²Ростов-на-Дону, Южный федеральный университет

При разработке стратегии социально-экономического развития южного региона ведется работа по оценке различных сценариев развития регионов и связанного с этим использования природно-ресурсного потенциала. Для выполнения поставленной задачи осуществляется разработка методов построения и оценивания сценариев использования природных ресурсов с применением современных методов обработки информации. Одним из таких методов является построение геоинформационной модели. Большой объем используемой информации позволяет получить более точные данные, однако именно геоинформационная модель превращает сложный процесс оценки данных в наглядный процесс принятия решений. В ходе исследования разработана геоинформационная модель для оценки природно-ресурсного потенциала региона, поиска функциональных и логических закономерностей в накопленных данных, построены модели и правила, которые объясняют найденные закономерности. В этих разработках находят применение методы распознавания образов, классификации и кластеризации. В докладе рассматриваются методы, основанные на спектральных характеристиках мультивременных композитов космоснимков, построение карт на основе большого массива данных, нашедшие применение при решении задач анализа структуры землепользования, идентификации и оценки экологического состояния территорий, оценки степени аридности территории, степени техногенной нагрузки на природную среду.

- [1] *Сурков Ф. А. Архипова О. Е. Математический инструментарий, информационные технологии и компьютерные программы для поддержки принятия решений по обеспечению устойчивого развития региона // Экологическая стратегия развития прибрежных регионов: география, окружающая среда, население.*, Ростов-на-Дону: ЮНЦ РАН, 2015. — С. 187–197.

Representation of large data in problems of decision support for sustainable development of the region

Arkhipova Olga^{1,2}

arkhipova@ssc-ras.ru

*Surkov Feodor*²*

fasurkov@sfedu.ru

¹Rostov-on-Don, Institute of Arid zones SSC RAS

²Rostov-on-Don, Southern Federal University

In the Southern and North Caucasus federal districts, when developing a strategy for social and economic development, work is underway to create tools for assessing various scenarios for the development of regions and the associated use natural resource potential, the state of ecological comfort of the region. To accomplish this task, we develop methods for constructing and evaluating scenarios for the use of natural resources using modern information processing methods. One of these methods is the construction of a geoinformation model(GIS) of the process. The large amount of information used allows obtaining more accurate data, but it is the geoinformation model(GIS) transforms the complex process of evaluating data into a visual decision-making process. In the course of the research, a geoinformation model was developed to assess the region's natural and resource potential, assess the impact of economic activity in the region on the state of natural resources. In these developments, methods of pattern recognition, classification and clustering, and methods of the construction of maps based on the big data are used. The report examines the tasks and methods that have found application in solving the following problems, analysis of the structure of land use based on satellite images, identification and assessment of the ecological state of territories, estimates of the degree of aridity of the territory, assessing the degree of man-caused environmental stress, using indices based on the spectral characteristics of multivariate composites

- [1] *Surkov F. A. Arkhipova O. E.* Mathematical tools, information technologies and computer programs to support decision-making for sustainable development of the region. *Ecological strategy for the development of coastal regions: geography, environment, population. Medico-ecological and socio-economic problems of coastal regions.* 1(1):187–197.

Алгоритм обработки изображений на основе диагональной аппроксимации графа смежности элементов изображения

Грачева Инесса Александровна^{1*}

gia1509@mail.ru

*Копылов Андрей Валериевич*¹

andkopylov@gmail.com

¹Тула, Тульский государственный университет

В большинстве работ по анализу изображений, использующих графовые модели, в качестве графа смежности элементов рассматривается четырехсвязная прямоугольная решетка. Кликовое число подобного графа равно двум и оптимизационная задача, возникающая при использовании байесовского подхода и оценки максимума апостериорной вероятности, является задачей парно-сепарабельной оптимизации. Но отсутствие учета диагональных связей между соседними пикселями приводит к снижению точности обработки. В то же время, прямое использование восьмисвязного графа смежности с диагональными связями приводит к резкому увеличению вычислительной сложности задачи.

Предлагаемое в рамках работы [1] расширение разработанной ранее гамма-нормальной модели и способа аппроксимации решетчатых графов смежности позволяет для учета диагональных связей между элементами изображения находить комбинированные оценки скрытого поля по леводиагональным графам, в которых все связи в столбцах, кроме одного, разорваны и по праводиагональным графам, в которых все связи в строках, кроме одной, разорваны. Вычислительная сложность алгоритма при этом остается линейной относительно количества элементов изображения.

Работа поддержана грантами РФФИ №№ 16-07-01039, 16-57-52042.

- [1] *Грачева И. А., Копылов А. В.* Алгоритм обработки изображений на основе диагональной аппроксимации графа смежности элементов изображения // Известия ТулГУ, Технические науки Тула: Издательство ТулГУ, 2017 (в печати).

An image processing algorithm based on the diagonal approximation of the adjacency graph of image elements

Gracheva Inessa^{1*}

author_gia1509@mail.ru

*Kopylov Andrey*¹

And.Kopylov@gmail.com

¹Tula, Tula State University

The majority of works on graph models based image analysis consider a four-connected rectangular lattice as the elements adjacency graph. The clique number of such a graph is equal to two and the optimization problem, arising when using the Bayesian approach and the maximum a posteriori probability estimation, is the pair-wise separable optimization task. But the lack of taking into account the diagonal relationships between neighboring pixels leads to a decreasing accuracy of processing. At the same time, direct usage of an eight-connected adjacency graph with diagonal constraints leads to a dramatic increase in the computational complexity of the problem.

The extension of the previously developed gamma normal model and the method of the lattice adjacency graphs approximation, proposed in the work [1] allows us to find combined estimates of the hidden field from the left-diagonal graphs in which all the links in the columns except one are broken and by right-diagonal graphs, in which all the links in the lines, except one, are broken. The computational complexity of the algorithm remains linear with respect to the number of pixels.

This research is funded by RFBR, grants 16-07-01039, 16-57-52042.

- [1] *Gracheva I. A., Kopylov A. V.* An image processing algorithm based on the diagonal approximation of the adjacency graph of image elements // Proceedings of TulSU, Tula: Tula State University, 2017 (in press).

Глубокие свёрточные автоэнкодеры: стереотождествление для восстановления 3D-моделей слабо текстурированных объектов

Князь Владимир Владимирович^{1,2} vl.kniaz@gosniias.ru
*Выголов Олег Вячеславович*¹ o.vygovlov@gosniias.ru
Федоренко Вадим Вадимович^{1*} vfedorenko@gosniias.ru
*Севрюков Владислав Дмитриевич*¹ vsevryukov@gosniias.ru

¹Москва, ФГУП «Государственный научно-исследовательский институт авиационных систем» (ФГУП «ГосНИИАС»)

²Москва, Московский физико-технический институт (ГУ)

Восстановление трёхмерных моделей объектов со слабо выраженными текстурами требует применения дескрипторов, способных разделять очень похожие друг на друга классы характерных элементов изображений. Дескрипторы особых точек, основанные на анализе градиентов (SIFT, ORB, LBP), не обеспечивают необходимого качества стереотождествления в случае слабо выраженных текстур, как, например, в случае археологических артефактов, только что извлечённых из грунта.

Предложен новый метод решения задачи стереотождествления на основе глубоких свёрточных автоэнкодеров. Автоэнкодер производит понижение размерности изображения на несколько порядков и формирует код, который может использоваться для решения задачи сопоставления соответственных точек.

Разработана архитектура автоэнкодера, выполняющего кодирование и восстановление цветных изображений, разрешением 32×32 пиксела. Приводится сравнение результатов работы предложенного метода и классических дескрипторов особых точек. Выполнена обработка снимков и восстановление 3D-моделей археологических раскопок, проводимых Босфорской экспедицией Государственного Исторического Музея.

Результаты применения разработанного метода показывают, что он превосходит традиционные дескрипторы особых точек на слабо текстурированных объектах и позволяет успешно решать задачу их 3D реконструкции.

Работа поддержана грантом РФФИ № 17-29-04410.

- [1] *Князь В. В., Выголов О. В., Федоренко В. В., Севрюков В. Д.* Глубокие свёрточные автоэнкодеры: стереотождествление для восстановления 3D-моделей слабо текстурированных объектов // Машинное обучение и анализ данных, 2017.

Deep convolutional autoencoders: stereo matching for 3D model reconstruction of low-textured objects

Kniaz Vladimir^{1,2}

vl.kniaz@gosniias.ru

*Oleg Viacheslavovich*¹

o.vygalov@gosniias.ru

*Fedorenko Vadim*¹*

vfedorenko@gosniias.ru

*Sevrykov Vladislav*¹

vsevryukov@gosniias.ru

¹Moscow, The Federal State Unitary Enterprise “State Research Institute of Aviation Systems” (FGUP “GosNIIAS”)

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

Powerful descriptors that can divide into groups very similar image patches are required to reconstruct 3D models of low-textured objects. Patch descriptors that are based on local gradients (SIFT, ORB, LBP) do not guarantee a robust stereo matching if the object has no bright textures, e.g., in the case of 3D model reconstruction of archaeological artifacts that were not cleaned from the soil.

We present a new method for stereo matching based on deep neural convolutional autoencoders. An autoencoder reduces the image dimensions and produces the code that could be used to perform an effective search of the corresponding image patch for low-textured object.

An architecture of a new autoencoder was developed. The autoencoder performs coding and decoding of color images with resolution 32×32 pixels. We present a comparison of the performance of the developed method and modern image patch descriptors. The method was applied to process images and reconstruct 3D models of archaeological excavations organized by the Bosphorus expedition of the Russian State Historical Museum.

The analysis of an application of the developed method proves that it outperforms the existing image descriptors in the matching of image patches of low-textured objects.

This research is funded by RFBR, grant 17-29-04410.

- [1] *Kniaz V. V., Vygalov O. V., Fedorenko V. V., Sevrykov V. D.* Deep convolutional autoencoders: stereo matching for 3D model reconstruction of low-textured objects // *Journal of Machine Learning and Data Analysis*, 2017.

Метод локализации областей интереса с текстурой специального вида на изображениях

Мурашов Дмитрий Михайлович^{1*} d_murashov@mail.ru
*Мурашов Федор Дмитриевич*² friedhelmsteuer@gmail.com

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

²Москва, Московский авиационный институт (НИУ)

В статье рассматривается задача локализации информативных областей с текстурой специального вида с элементами однородной пространственной ориентации и разными пространственными частотами. Задача локализации информативных областей решается как задача сегментации текстурных изображений. Существующие методы, на основе признаков Харалика, энергетических признаков Лавса и фильтров Габора не позволяют полностью решить задачу с требуемым качеством. Предлагается метод на основе модифицированного суперпиксельного алгоритма с процедурой постобработки. Вектор описания пикселей изображения расширен текстурными признаками, вычисляемыми по компонентам структурного тензора. Применение модифицированного алгоритма позволяет учесть пространственные, цветовые и текстурные свойства изображений. Для обеспечения приемлемого качества сегментации используется условие минимума меры избыточности информации. Проведен вычислительный эксперимент на тестовых текстурных изображениях. Результаты сегментации изображения текстурной мозаики предложенным методом сравнивались с результатами известного метода на основе энергетических признаков Лавса. Разработанный метод был применен для локализации информативных областей на изображениях произведений живописи. Результаты эксперимента показали эффективность предлагаемого метода.

Работа частично поддержана грантами РФФИ №15-07-09324 и №15-07-07516.

- [1] *Мурашов Д. М., Мурашов Ф. Д.* Метод локализации информативных областей с текстурой специального вида // Машинное обучение и анализ данных, 2017.

Method for localizing image regions of interest with a specific texture

Murashov Dmitry^{1*}

d_murashov@mail.ru

*Murashov Fedor*²

friedhelmsteuer@gmail.com

¹Moscow, Federal Research Center “Computer Science and Control”
of RAS

²Moscow, Moscow Aviation Institute (NRU)

The paper deals with a problem of localization of informative regions with a specific texture in digital images. This type of texture is characterized by uniformly oriented elongated elements and varying spatial frequency. Such a structure, in particular, can be generated by groups of brushstrokes in the images of paintings. Existing methods based on the Haralick’s features, Laws energy features and Gabor filters can not completely solve the problem with the required quality. In this paper, the task of localization of informative areas will be addressed as a problem of segmentation of texture images. A method for solving the problem based on a modified super-pixel segmentation algorithm with a postprocessing procedure is proposed. Vector of image pixel description is expanded by texture features computed using components of the structure tensor. The selected features involve the peculiarities of the considered texture type. Application of superpixel algorithm with an extended feature description of images will permit to take into account spatial, color, and textural properties of image regions. To obtain an acceptable quality of segmentation, the condition of minimum information redundancy measure is used. A computational experiment is carried out on textural test images. The results of segmenting image of the texture mosaic by the proposed method and the well-known method based on the Laws’s energy features, are compared. Comparison demonstrates the advantage of the proposed method. The developed technique is used to localize informative areas in the images of paintings. The results of the experiment show the efficiency of the proposed method.

This research is funded partially by RFBR, grants 15-07-09324 and 15-07-07516.

- [1] *Murashov D., Murashov F.* Method for localizing informative regions with texture of a special type // *Journal of Machine Learning and Data Analysis*, 2017.

Поиск фоновых областей в изображении

Нестеренко Виктор Александрович

neva09@mail.ru

Ростов-на-Дону, Институт математики, механики и компьютерных наук ЮФУ

Первым этапом задачи распознавания в большинстве случаев является локализация объекта на изображении: выделение связанной группы пикселей по каким либо признакам отличающихся от остальной части изображения. В том случае, если характеристики объекта заранее не известны (например поиск объектов искусственного происхождения на аэроснимках), к решению задачи локализации можно подойти с другой стороны: вначале определить, что является фоном изображения и считать участки изображения отличные от фона искомыми объектами.

В данном докладе рассматривается метод выявления фоновых областей изображения основанный на выявлении сходных по характеристикам, связанных групп пикселей. В качестве характеристик используются признаки Хаара с соответствующим набором масок. Для выявления близких по характеристикам групп пикселей применяется FCM (Fuzzy Classifier Means) метод кластеризации данных.

- [1] *Нестеренко В. А.* Поиск фоновых областей в изображении // Информационные процессы, 2016, т. 16, в. 3, С. 291–300. <http://www.jip.ru/2016/291-300-2016.pdf>.

The search for background areas in an image

Nesterenko Victor

neva09@mail.ru

Rostov-on-Don, Institute of Mathematics, Mechanics and Computer
Sciences of SFedU

The first step of a task of recognition in most cases is localization of object on the image: allocation of linked group of pixels different from other part of the image. When the characteristics of object aren't known in advance (the search for the objects of artificial origin on aeropictures for example), the solution of the problem of localization it is possible to approach from the other side: at first to determine, what is the background of image and to consider the sections of image other than a background as required objects.

In this report the method of identification of background areas of the image based on identification according to characteristics linked groups of pixels is considered. The signs of Haar with the corresponding set of masks are used as descriptions of pixels. The FCM (Fuzzy Classifier Means) method of a clustering of data is applied to identification of relatives according to characteristics of groups of pixels.

- [1] *Nesterenko V. A.* The search for background areas in an image // Information processes, 2016, v. 16, no. 3, P. 291–300. <http://www.jip.ru/2016/291-300-2016.pdf>.

Построение цифровых моделей рельефа по растровым топографическим картам

*Титов Виталий Семенович*¹

titov-kstu@rambler.ru

Мирошниченко Сергей Юрьевич^{1*}

oldguy7@rambler.ru

¹Курск, Юго-Западный государственный университет

Разработана методика построения цифровых моделей рельефа (ЦМР) по топографическим картам, которая за счет применения методов обработки и распознавания изображений обеспечивает автоматический режим функционирования.

Рассмотрены основные элементы представления рельефа на топографических картах, показаны особенности их графического отображения. Представлен метод формирования множества векторных представлений горизонталей, отличающийся возможностью локализации их центров с учетом наличия микрорельефа и позволяющий корректно выделять близкорасположенные изолинии.

Описан метод выделения отметок высот и подписей горизонталей, обеспечивающий инвариантность распознавания к размерам и углу поворота составляющих их цифр. Рассмотрены методы присвоения высот выделенным горизонталям и контроля точности векторной ЦМР с помощью анализа графа значений высот.

В результате экспериментальной проверки предложенной методики установлено, что длительность автоматического построения ЦМР для одного номенклатурного листа карты, сканированной в разрешении 400 точек/дюйм, независимо от масштаба составляет не более 10 мин и варьируется в соответствии с количеством элементов рельефа.

- [1] *Титов В. С.* Построение цифровых моделей рельефа по растровым топографическим картам // Телекоммуникации, Москва: Наука и технологии, 2017. — С. 34–43.

Digital elevation models creation from raster topographic maps

*Titov Vitaliy*¹

titov-kstu@rambler.ru

Miroshnichenko Sergey^{1*}

oldguy7@rambler.ru

¹Kursk, Southwest State University

A technique for constructing digital elevation models (DEM) based on topographic maps has been developed, which, due to the application of image processing and recognition methods, provides an automatic mode of operation.

Main relief elements on topographic maps and their visual representation features are described. The method for vector isolines' set extraction, that differs from analogs by lines' centers detection ability and allows to separate nearly placed horizontals is described.

The presented method for height values detection on maps is based on contour analysis and provides rotation and scale invariant digits recognition. The methods for height values assignment to vector isolines and error control uses height vertex graph analysis.

Experimental verification of developed methodic shows duration of that automatic DEM creation from a map's single standard sheet (scanned in 400 dpi resolution) does not exceed 10 minutes and varies depending on relief elements quantity.

- [1] *Titov V. S.* Digital elevation models creation from raster topographic maps // Telecommunications, Moscow: Science and Technology, 2017. — p. 34–43.

Модель применения метода Уорда для кластеризации пикселей изображения

Ханыков Игорь Георгиевич¹

igk@iias.spb.su

Харинов Михаил Вячеславович^{1*}

khar@iias.spb.su

¹ Санкт-Петербург, Институт информатики и автоматизации РАН

Рассматривается иерархическая кластеризация пикселей для детектирования объектов [1]. Детектирование сводится к построению иерархической последовательности кусочно-постоянных приближений изображения, в которых пиксели различаются по цвету и описывают объекты, выделяемые компьютером. Основной проблемой *унифицированного* детектирования объектов на изображениях произвольного содержания является устойчивость выделения объектов на различных изображениях сцены. Если единообразие выделяемых компьютером объектов заметно нарушается, то детектирование считается *неустойчивым*. При всем многообразии современных методов машинного детектирования объектов лишь немногие из них являются устойчивыми. Детектирование объектов методом Уорда иерархической кластеризации пикселей является одним из устойчивых методов, который недостаточно применяется в практике обработки изображений из-за чрезмерной вычислительной сложности.

Для применения метода Уорда к изображениям из миллионов пикселей развивается модель изображения, учитывающая особенности входных данных, и модель вычислений, в которой формализуются способы иерархического представления, упорядочения и преобразования кластеров пикселей в компьютере в терминах алгебраической сети, формируемой посредством слияния сопоставляемых пикселям элементарных сетей, где под элементарной сетью понимается дерево (ациклический граф) и цикл (циклический граф), содержащие по одному узлу.

Структура данных, конкретные алгоритмы и экспериментальные результаты представлены в докладе.

- [1] Ханыков И. Г., Харинов М. В. Модель цифрового изображения на основе модифицированного метода Уорда кластеризации пикселей // Вестник Бурятского государственного университета. Математика, информатика, Улан-Уде, 2017. № 2 — С. 61–70.

A model for implementation of Ward's method of image pixel clustering

*Khanykov Igor*¹

igk@iiias.spb.su

*Kharinov Mikhail*¹★

khar@iiias.spb.su

¹Saint-Petersburg, Saint-Petersburg Institute for Informatics and Automation of RAS

The hierarchical pixel clustering for the detection of the objects is considered [1]. The detection is reduced to construction of hierarchical sequence of piecewise constant approximations of the image. According to the model pixels of distinguished colors describe objects that are detected by a computer. The main problem of *unified* detection of objects in the images of arbitrary content is the stability of object detection in various images of the scene. If the consistency of the object detection by computer is noticeably violated, then the object detection is considered to be *unstable*. Only a few of all the variety of modern methods of the machinery object detection are stable. The object detection by the Ward method of hierarchical pixel clustering is one of the stable methods. Alas, it is rarely applied in image processing practice due to excessive computational complexity.

To apply the Ward's method for the images consisting of millions of pixels, we develop the image model that takes into account the characteristics of input data and the computation model. In the latter the methods of hierarchical representation, ordering and transformation of pixel clusters in RAM are formalized in terms of an algebraic network formed by merging pixel-matched elementary networks. The elementary network designates a tree (an acyclic graph) and a cycle (a cyclic graph), containing each a single node.

The data structure, specific algorithms and experimental results are presented in the report.

- [1] *Khanykov I., Kharinov M.* A model of digital image preprocessing based on pixel clustering by modified Ward's method // Bulletin of the Buryat State University. Mathematics, Informatics, Ulan-Ude, 2017. No 2 — p. 61–70.

Влияние свойств данных дистанционного зондирования земли и параметров их трансформирования на время построения результирующего изображения местности

*Шуклин Игорь Игоревич*¹

shuklin@nm.ru

Мосин Сергей Александрович^{2*}

22mosin@gmail.com

*Мирошниченко Сергей Юрьевич*²

oldguy7@rambler.ru

¹Курск, ФГУП «18 ЦНИИ» МО РФ

²Курск, Юго-Западный государственный университет

Авторами проведены и описаны экспериментальные исследования влияния свойств данных дистанционного зондирования Земли и параметров их трансформирования на время построения результирующего изображения. Для множества свойств исходных данных дистанционного зондирования и соответствующих трансформированных изображений местности выделены такие свойства, как формат хранения исходных растровых данных и их внутренняя структура. Среди множества используемых методов и параметров трансформирования изображений местности выделены типы полиномов трансформации и алгоритмов сглаживания. Рассмотрены такие параметры аппаратной платформы, как тип дискового массива и размер кластера файловой системы. Показано влияние степени сжатия файлов на скорость выполнения трансформирования, рассмотрены изменения в скорости при смене формата (формат выходного изображения отличается от формата входного изображения).

По результатам исследований были вынесены рекомендации по заданию параметров трансформирования космических изображений местности и настройке цифровых фотограмметрических станций.

- [1] Шуклин И. И., Мосин С. А., Мирошниченко С. Ю. Влияние свойств данных дистанционного зондирования земли и параметров их трансформирования на время построения результирующего изображения местности // Известия Юго-Западного государственного университета, Курск: Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Юго-Западный государственный университет», 2016. — С. 16–27.

Earth remote sensing data features and transformation parameters effect on terrain image creation process duration

*Shuklin Igor*¹

shuklin@nm.ru

Mosin Sergey^{2*}

22mosin@gmail.com

*Miroshnichenko Sergey*²

oldguy7@rambler.ru

¹Kursk, FGUP "18 TsNII" MO RF

²Kursk, Southwest State University

This article reports experimental results of remote sensing data processing speed increase. The authors led and described an effect of remote sensing data properties and their transformation's parameters on terrain image creation process duration. The research is actual due to the need of digital photogrammetric stations performance increase. The authors defined the properties of the initial data: raster data storage format and internal structure. Similarly, authors defined the properties of the transformed data.

Among the many transformation methods and transformation parameters were selected the type of polynomial transformation and smoothing algorithm. Hardware platform options were defined as following: the storage array type and the file system cluster size. The authors have imposed restrictions on the hardware platform, methods of image transformation, dimensions of the input and output pixel image, coordinate system. The authors have shown the effect of image files compression ratio on the speed of the transformation. The authors considered the changes in transformation process speed when changing the format.

Based on studies made recommendations for the selection of transformation parameters values that will help to improve digital photogrammetric stations performance without accuracy and visual properties loss.

- [1] *Shuklin I. I., Mosin S. A., Miroshnichenko S. Yu.* Earth remote sensing data features and transformation parameters effect on terrain image creation process duration // Proceedings of the South-West State University, Kursk: Federal State Budgetary Educational Institution of Higher Education "Southwest State University", 2016. — p. 16–27.

Разработка метода защиты от высококачественных подделок для мобильной системы распознавания по радужной оболочке

Ефимов Юрий Сергеевич^{1*}

yuri.efimov@phystech.edu

*Одиноких Глеб Андреевич*²

g.odinokikh@gmail.com

*Соломатин Иван Андреевич*¹

ivansolomat@yandex.ru

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Федеральный исследовательский центр «Информатика и управление» РАН

Одной из проблем современных систем распознавания по радужной оболочке является их уязвимость по отношению к подделкам данного биометрического признака. Мобильные биометрические системы более подвержены атакам при помощи подделок вследствие ограниченных ресурсов и меньшего интервала времени.

Рассматриваются статические способы определения живости глаза, представленного биометрической системе, поскольку мобильные системы распознавания подразумевают высокую степень быстродействия и зачастую используют распознавание по единственному кадру. Проанализировано два подхода: на основании анализа текстурных особенностей изображений подделок радужной оболочки и применения глубоких нейронных сетей для выделения характерных особенностей таких изображений.

Помимо описанных в литературе способов атаки, рассмотрены два новых способа подделки при помощи распечаток изображений человеческих глаз в высоком разрешении. При помощи инфракрасной камеры собрана и размечена соответствующая база изображений.

Проведён вычислительный эксперимент для проверки работоспособности на открытых базах изображений и собранной вручную базе. Точность определения живости представленного на изображении глаза сравнима с таковой у описанных в современной литературе аналогов.

Работа поддержана грантом РФФИ № 16-07-01171.

- [1] *Ефимов Ю. С., Одиноких Г. А., Соломатин И. А.* Разработка метода защиты от высококачественных подделок для мобильной системы распознавания по радужной оболочке // Машинное обучение и анализ данных, 2017.

High-quality presentation attack detection in a mobile iris recognition system

*Efimov Yuriy*¹★

yuri.efimov@phystech.edu

*Odinokikh Gleb*²

g.odinokikh@gmail.com

*Solomatina Ivan*¹

ivansolomat@yandex.ru

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

²Moscow, Federal Research Center “Computer Science and Control”
of RAS

Modern iris authentication systems are vulnerable to presentation attacks by fake self-manufactured synthetic samples, which requires the development of new and efficient protection measures. Mobile iris recognition systems are more prone to *spoofing* because of user interaction difficulties and performance limitations.

Software-based static methods of iris liveness detection are proposed, since mobile systems require fast interaction for user convenience and basically single frame recognition is used. Two different approaches are proposed: iris texture analysis using local descriptors and using a deep CNN to classify iris images.

Apart from spoofing methods by printed infra-red eye images or contact lens, which are well described in modern literature, a new type of a presentation attack is proposed and the corresponding image database is collected using an infra-red camera.

Computation experiment was performed on public domain databases and on a collected database. Liveness detection accuracy is comparable to that of state-of-the-art methods and outperforms many of them.

This research is funded by RFBR, grant 16-07-01171.

- [1] *Efimov Yu., Odinokikh G., Solomatina I.* Detection of high-quality presentation attacks for a mobile iris recognition system // Journal of Machine Learning and Data Analysis, 2017.

Комплексирование данных для классификации объектов по ансамблю изображений

Ланге Михаил Михайлович^{1*}
*Ганебных Сергей Николаевич*¹

lange_mm@ccas.ru
 sng@ccas.ru

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

Исследуются метрические многоклассовые классификаторы на ансамбле изображений от источников различной модальности. Классификаторы принимают коллективные решения по составным объектам, заданным наборами изображений, по одному от каждого источника. Разделяющие функции порождаются элементарными NN или SVM классификаторами типа «класс против всех». Исследуются две оригинальные схемы комплексирования данных, которые используют в разделяющих функциях либо композиции мер на множествах изображений источников (General Measure), либо композиции мягких решений по изображениям источников (General Similarity). В терминах вероятности ошибок, предлагаемые схемы сравниваются с известной схемой голосования композиций жестких решений по изображениям источников (Majority Vote). Экспериментальные оценки вероятности ошибок распознавания лиц по декоррелированным компонентам I_1 , I_2 , I_3 цветного RGB изображения и по ансамблю (I_1 , I_2 , I_3) с использованием схем MV, GM и GS даны в таблице и демонстрируют преимущество схем GM и GS по сравнению со схемой MV.

	компоненты			схема комплексирования		
	I_1	I_2	I_3	MV	GS	GM
NN	0,0034668	0,0005744	0,0012992	0,0000048	0,0000012	0,0000020
SVM	0,0019128	0,0023692	0,0016508	0,0001772	0,0000072	0,0000008

Полная версия статьи дана в [1].

Работа поддержана грантом РФФИ 15-01-04671, 15-07-09324.

- [1] *Ганебных С. Н., Ланге М. М.* Об эффективности схем комплексирования данных для распознавания образов по ансамблю изображений // Машинное обучение и анализ данных, 2017.

Data Fusion Schemes for Object Classification Using Ensemble of Images

*Lange Mikhail*¹*

lange_mm@ccas.ru

*Ganebnykh Sergey*¹

sng@ccas.ru

¹Moscow, Federal Research Center “Computer Science and Control” of RAS

In an ensemble of image sources of different modalities, some metric multiclass classifiers are studied. The classifiers make the collective decisions for the composite objects that are produced by collections of the images with one from each source. The discriminant functions of the multiclass classifiers are produced by the binary “class-vs-all” NN or SVM classifiers. Two original fusion schemes that use the discriminant functions based on the different compositions are suggested. The first scheme uses the compositions of the dissimilarity measures between the images within each source (General Measure) whereas the second scheme uses the compositions of the soft decisions for the images in the submitted composite object (General Similarity). In terms of error rates, the proposed GM and GS fusion schemes are compared with the known MV scheme which is based on majority voting the compositions of the hard decisions for the individual source images. The table contains the experimental error rates for face recognition using the individual decorrelated components I_1 , I_2 , I_3 of the color RGB image as well as the MV, GM, and GS fusion schemes in the ensemble (I_1 , I_2 , I_3). The results show an advantage of GM and GS schemes with respect to MV scheme. The full version of the paper is given in [1].

	компоненты			схема комплексирования		
	I_1	I_2	I_3	MV	GS	GM
NN	0,0034668	0,0005744	0,0012992	0,0000048	0,0000012	0,0000020
SVM	0,0019128	0,0023692	0,0016508	0,0001772	0,0000072	0,0000008

This research is funded by RFBR, grant 15-01-04671, 15-07-09324.

- [1] *Ganebnykh S., Lange M* On Efficiency of Fusion Schemes for Pattern Recognition in Ensemble of Images // Journal of Machine Learning and Data Analysis, 2017.

Полноконволюционная сеть FaceDetectNet для обнаружения лиц

*Моисеевко Анастасия Сергеевна*¹★ ms.asmoiseenko@gmail.com
*Горбачевич Владимир Сергеевич*¹ gvs@gosniias.ru
*Визильтер Юрий Валентинович*¹ viz@gosniias.ru

¹Москва, ФГУП «Государственный научно-исследовательский институт авиационных систем» (ФГУП «ГосНИИАС»)

Обнаружение лиц является одной из самых популярных задач компьютерного зрения. Задача обнаружения лиц подразумевает поиск и локализацию лиц на изображениях для дальнейшей автоматической обработки изображений: распознавание лиц, трехмерное моделирование лиц и т. д. Существует множество подходов к решению задачи обнаружения лиц, включая различные алгоритмы на основе глубоких конволюционных сетей (CNN), однако проблема оптимального соотношения между качеством обнаружения и скоростью вычислений по-прежнему остается актуальной. В данной статье мы предлагаем новое решение задачи обнаружения лиц с использованием CNN под названием FaceDetectNet. Наша архитектура CNN основана на идеях архитектуры YOLO/DetectNet и GoogleNet с некоторыми новыми деталями реализации. Мы предлагаем оригинальный алгоритм итеративной кластеризации ограничивающих прямоугольников для агрегирования выходов CNN. Наш алгоритм FaceDetectNet обучается и тестируется на наиболее сложной базе изображений WiderFace и достигает средней точности 0,69 на наборе данных WiderFace/hard и, таким образом, превосходит все конкурирующие детекторы на сложной подвыборке, помимо современного решения HR, основанного на существенно более глубокой и медленной CNN, в то время как наш алгоритм FaceDetectNet может работать в режиме реального времени на графическом процессоре NVIDIA GeForce 1080 GPU. Таким образом, наш подход обеспечивает лучшее качество с разумной вычислительной скоростью.

Работа выполнена при финансовой поддержке РФФ, грант № 16-11-00082.

[1] *Моисеевко А. С., Горбачевич В. С., Визильтер Ю. В.* Полноконволюционная сеть FaceDetectNet для обнаружения лиц // Компьютерная оптика, 2017 (в печати).

FaceDetectNet Fully-convolutional Network for Face Detection

*Moiseenko Anastasiia*¹*

ms.asmoiseenko@gmail.com

*Vizilter Yury*¹

viz@gosniias.ru

*Gorbatsevich Vladimir*¹

gvs@gosniias.ru

¹Moscow, The Federal State Unitary Enterprise “State Research Institute of Aviation Systems” (FGUP “GosNIIAS”)

Face detection is one of the most popular computer vision tasks. There are a lot of face detection approaches proposed including different CNN-based techniques, but the problem of optimal balancing between detection quality and computational speed is still relevant. In this paper we propose new CNN-based solution for face detection called FaceDetectNet. Our CNN architecture is based on ideas of YOLO/DetectNet and GoogleNet architecture supported with some new tools and implementation details created especially for our face detection application. We propose: original iterative proposal clustering (IPC) algorithm for aggregation of output face proposals formed by CNN and the 2-level “weak pyramid” providing better detection quality on the testing sets containing both small and huge images. Our face detection approach is close to previously proposed SSD-based face detection, but the principal difference is that we use the deep features of top hidden CNN layer for forming the face proposals of any size. Thus we utilize the global semantic and context information for improving the detection quality for small faces. Our FaceDetectNet is trained and tested on the most challenging WIDER FACE detection benchmark. Our algorithm achieves the average precision (AP) 0.69 on the WIDER FACE hard level, and thus outperforms all competitive detectors on the Hard level besides the HR state-of-the-art solution. Note that HR solution is based on essentially deeper and slower CNN, while our FaceDetectNet can work in real-time on the NVIDIA GeForce 1080 GPU. On the other hand, SSD-based face detector with comparable CNN parameters provides AP 0.625 only on the WIDER FACE hard level. So, our approach provides the best quality with reasonable computational speed.

This research is funded by RSF, grant 16-11-00082.

- [1] *Moiseenko A., Gorbatsevich V. and Vizilter Y.* FaceDetectNet Fully-convolutional Network for Face Detection // Computer Optics, 2017 (in press).

Распознавание изображений на основе метода максимально правдоподобных рассогласований

Савченко Андрей Владимирович

avsavchenko@hse.ru

Нижегород, Национальный исследовательский университет
«Высшая школа экономики»

В [1] исследуется задача распознавания изображений в условиях малых выборок наблюдений на основе технологии transfer learning. Предложен алгоритм распознавания на основе максимизации совместной плотности вероятности (правдоподобия) рассогласований между признаками входного изображения и всех эталонов, извлеченных на выходе одного из последних слоев сверточной нейронной сети с функцией активации ReLU. Для оценки правдоподобия используется известное асимптотически нормальное распределение рассогласования Йенсена-Шеннона между векторами высокой размерности. На предварительном этапе предлагаемого метода вычисляется матрица попарных рассогласований между классами. В процессе распознавания определяется степень близости входного изображения со всеми эталонами, и для каждого класса оценивается логарифм правдоподобия вычисленных рассогласований. Показано, что предложенный подход может применяться не только совместно с дивергенцией Йенсена-Шеннона, но и для произвольной меры близости. Приведены результаты экспериментального исследования для задач распознавания лиц с признаками, выделяемыми с помощью нейросетевых моделей VGGFace и Lighened CNN (B). Для наборов данных Labeled Faces in the Wild и YouTube Faces показано, что предлагаемый алгоритм позволяет на 1-5% повысить точность метода ближайшего соседа для классификации изображений и видео.

Исследование выполнено при поддержке гранта президента РФ для молодых ученых – докторов наук № МД-306.2017 и Лаборатории алгоритмов и технологий анализа сетевых структур Национального исследовательского университета Высшая школа экономики.

- [1] *Савченко А. В.* Метод максимально правдоподобных рассогласований в задаче распознавания изображений на основе глубоких нейронных сетей // Компьютерная оптика, 2017. — Т.41, № 3. — С. 422–430.

Image recognition based on the method of maximum likelihood dissimilarities

Savchenko Andrey

avsavchenko@hse.ru

Nizhny Novgorod, National Research University Higher School
of Economics

In [1] we explore the image recognition problem in the case of small sample size based on the transfer learning techniques. We propose the recognition algorithm based on the maximum likelihood (joint density) of dissimilarities between an observed image and available instances in the training set. The feature vectors are extracted at the output of one of the last layers of the convolutional neural network with ReLU activations. The likelihood is estimated using the known asymptotically normal distribution of the Jensen-Shannon divergence between high-dimensional vectors. The matrix of pairwise distances between all classes is computed at preliminary stage. During recognition procedure, an observed image is matched with all instances from the training set. After that, the negative log-likelihood of the computed dissimilarities is estimated. It is shown that the proposed approach can be applied with an arbitrary dissimilarity measure rather than with the Jensen-Shannon divergence, which asymptotic properties are used to estimate the unknown likelihood. The experimental study is focused on unconstrained face recognition with features extracted by VGGFace and Lightened CNN (B) neural network models. We demonstrate that the proposed approach makes it possible to increase the recognition accuracy of the nearest neighbor rule for classifying images and videos at 1-5%.

The work is supported by Russian Federation President grant no. MD-306.2017.9 and Laboratory of Algorithms and Technologies for Network Analysis, National Research University Higher School of Economics.

- [1] *Savchenko A.* Maximum-likelihood dissimilarities in image recognition with the deep neural networks // *Computer Optics*, 2017. — Vol.41, No. 3. — p. 422–430.

Дескриптор в аккумуляторном пространстве Хафа градиентного поля изображения для детектирования пешеходов

Самсонов Никита Андреевич^{1*} nikita.samsonov@phystech.edu

Гнеушев Александр Николаевич² gneushev@ccas.ru

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Федеральный исследовательский центр «Информатика и управление» РАН

Автоматическое детектирование объектов на изображениях и видео является одной из основных задач компьютерного зрения. Такие объекты на изображении, как люди, являются неотъемлемой частью дорожной сцены, играют важную роль в развитии дорожной ситуации. Сложность распознавания человека на изображении обусловлена неравномерной освещенностью, большой вариабельностью изображений человека из-за разных ракурсов, частичными перекрытиями человека другими объектами.

В данной работе предлагается алгоритм выделения локальных текстурных признаков на основе Гистограмм Аккумуляторного пространства Хафа (НАН), обобщающий НОГ дескриптор путем получения проекций не только ориентаций но и положений локальных контуров объекта на изображении. Дескриптор НАН строится на основе проекций в аккумуляторном пространстве Хафа, полученного с помощью лучевого поперечного преобразования Радона градиентного поля исходного изображения. НАН дескриптор учитывает не только локальное распределение ориентаций контурных признаков, но так же и их положение. Таким образом, предложенный метод преодолевает недостатки НОГ, используя дополнительной информации о положении локальных особенностей объекта.

Предлагаемый метод был протестирован совместно с линейным классификатором SVM на тестовой выборке пешеходов INRIA с результатом $miss\ rate = 0.039$ при $FPPW = 10^{-4}$, что в два раза лучше чем у НОГ ($miss\ rate = 0.079$).

Работа подддержана грантом РФФИ № 16-07-01171.

- [1] Самсонов Н. А., Гнеушев А. Н. Дескриптор в аккумуляторном пространстве Хафа градиентного поля изображения для детектирования пешеходов // Машинное обучение и анализ данных, 2017.

Textural descriptor in the Hough accumulator space of the gradient field for detecting pedestrians

Samsonov Nikita^{1*}

nikita.samsonov@phystech.edu

*Gneushev Alexander*²

gneushev@ccas.ru

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

²Moscow, Federal Research Center “Computer Science and Control”
of RAS

Automatic detection of objects in images and video is one of the main tasks of computer vision. Detecting humans in the image is an important task for road scene analyzing. There are some complexities of human detection in the image due to irregular illumination, large variability of human images because of pose and variable angles, partial overlapping of a person by other objects.

The image of a person can be represented by a set of contours, silhouettes of parts of the body which are appeared in the image as set of edges. However, the contours in the image of a person can have a complex structure, it is necessary to use a set of local descriptors for they characterization. This approach for the task of detecting may be used in an effective scheme with constructing an integral characteristic vector from local contour descriptors – Histograms of Oriented Gradients (HOG) in image areas. However, HOG does not fully describe the local contour features of the object and in real applications this method has an unacceptable number of false positives.

Generalization of the HOG descriptor by obtaining distributions not only of the orientations, but also of the positions of the edges in the local area of the image - Hough Accumulator Histograms (HAH) is proposed. The HAH descriptor is constructed on the basis of the histograms of the Hough accumulator space, obtained by means of the Radon transform of the image gradient field.

The proposed method was tested together with the SVM linear classifier on the pedestrian image test database INRIA with the result *Miss Rate* equal to 0.039 with *FPPW* equal to 10⁻⁴, which is twice as good as for HOG (*miss rate* = 0.079).

This research is funded by RFBR, grant 16-07-01171.

- [1] *Samsonov N., Gneushev A.* Textural descriptor in the Hough accumulator space of the gradient field for detecting pedestrians // *Journal of Machine Learning and Data Analysis*, 2017.

Исследование быстродействия распознавания 3D изображения методами стохастической геометрии в зависимости от количества опорных точек на сфере

*Федотов Николай Гаврилович*¹ fedotov@pnzgu.ru
Сёмов Алексей Александрович^{2*} matematik_aleksey@mail.ru
*Моисеев Александр Владимирович*¹ moigus@mail.ru

¹Пенза, Пензенский государственный университет

²Пенза, ООО «Комэрф»

В статье рассматривается новый подход к распознаванию 3D изображений, дающий инвариантное описание объекта при любой его пространственной ориентации. Метод позволяет параллельно с распознаванием производить анализ исходного изображения, а также конструировать признаки с заданными свойствами.

Предлагаемый подход основан на методах стохастической геометрии и функционального анализа. Разработанное авторами гипертрейс-преобразование создаёт 3D трейс-образ исходного пространственного объекта, который является удобным инструментом для анализа 3D изображения в отличие от математического аппарата других методов.

Стохастическое сканирование со случайными параметрами более эффективно по сравнению с детерминированным сканированием с точки зрения соотношения «надежность-быстродействие» распознавания 3D изображений. Приведены результаты экспериментов, показывающие не только теоретическую, но и практическую значимость, и эффективность предлагаемого подхода.

В статье анализируется быстродействие распознавания 3D объектов в зависимости от количества опорных точек на сфере с использованием различных видов сканирования. Предлагаются возможные пути по ускорению работы распознающей системы.

Работа поддержана грантом РФФИ № 15-07-04484.

- [1] *Федотов Н. Г., Сёмов А. А., Моисеев А. В.* Исследование быстродействия распознавания 3D изображения методами стохастической геометрии в зависимости от количества опорных точек на сфере // Машинное обучение и анализ данных, 2017. <http://jmla.org/papers/doc/2017/no1/Fedotov2017Research.pdf>.

Performance investigation of 3D image recognition by stochastic geometry methods in dependent on the number of reference points on the sphere

*Fedotov Nikolai*¹

fedotov@pnzgu.ru

Syemov Alexey^{2*}

matematik_aleksey@mail.ru

*Moiseev Alexander*¹

moigus@mail.ru

¹Penza, Penza State University

²Penza, LLC Comearth

In this article a new developed approach to the 3D images recognition, giving the object invariant description for any its spatial orientations is proposed. This method has many advantages and 3D images data mining capabilities. In particular, in parallel with the spatial object recognition it is possible to analyze the original image. Due to building a rigorous mathematical model, it is possible to design analytically features with predetermined properties.

The suggested approach is based on modern methods of stochastic geometry and functional analysis. Hypertrace transform creates a 3D trace-image of the original spatial object due to scan the parallel planes grid from different view angles. Created on this trace-image basis hypertrace matrix is a convenient tool for analyzing 3D images in contrast to other methods mathematical tool.

Stochastic scan with random parameters is more efficient than determinate scan in terms of the 3D images recognition "reliability - performance" relation. The conducted experiments results are showed. These results demonstrate both the theoretical and practical significance and effectiveness of the proposed method.

In the article the evaluation task of 3D image recognition performance in dependent on the number of reference points on the sphere with the use of scan various kinds are analyzed. At the article end potential further ways to accelerate the recognition system work are proposed.

This research is funded by RFBR, grant 15-07-04484.

- [1] *Fedotov N.* Performance investigation of 3D image recognition by stochastic geometry methods in dependent on the number of reference points on the sphere // Journal of Machine Learning and Data Analysis, 2017. <http://jmla.org/papers/doc/2017/no1/Fedotov2017Research.pdf>.

Метод биометрической идентификации по изображению внешней стороны ладони

Чернышов Виктор Геннадьевич

webcreator18@gmail.com

Москва, МГУ им. М. В. Ломоносова, Факультет ВМК

Наряду с сетчаткой глаза и отпечатками пальцев, использование внешней стороны ладони – одно из наиболее перспективных направлений в актуальных прикладных задачах, связанных с распознаванием личности человека. Ранее автором была представлена мобильная система биометрической идентификации, реализованная на базе приложения под ОС Android. В данной статье предложен эффективный метод генерации и сопоставления признаков, в ходе тестирования которого на реальных данных (42 человека, 486 изображений) для задачи идентификации был достигнут EER (равный уровень ошибок) = 0.5%.

Основным инструментом для формирования признаков выбран двумерный фильтр Габора, структура которого позволяет извлекать информацию о характере кожных складок пальцев на изображении: толщине, направлении, взаимном расположении.

Различные типы текстурных особенностей имеют разные характерные размеры. Поэтому, во-первых, введено два семейства фильтров с разными параметрами σ (масштаб кодируемых особенностей) и λ (пропорционально «периоду» квазиопериодических структур кожных складок). Семейством будем называть множество $\{G_R(x, y, \theta_j, \dots), j = 0, 1, \dots, 5\}$, где $G_R(\dots)$ – ядро вещественной части фильтра Габора с фиксированными параметрами σ, γ, λ ; θ_j – угол поворота фильтра, $\theta_j = j\pi/6$.

Во-вторых, для получения вектора признаков использовалось не полное изображение, а несколько его точек в узлах сетки, шаг которой по обоим направлениям пропорционален соответствующим размерам окна фильтра. Эксперименты показали, что такая модификация не ухудшает качество классификации, однако существенно уменьшает время сравнения пары признаков.

- [1] *Chernyshov V.* Efficient Hand Detection on Client-server Recognition System // Proceedings of the 10th International Conference on Computer Vision Theory and Applications - Volume 2, Berlin: INSTICC, 2015. – p. 461–468. <http://www.scitepress.org/DigitalLibrary/PublicationsDetail.aspx?ID=PzA/cNUBpXE=&t=1>.

Biometric identification using hand dorsal surface

Chernyshov Victor

webcreator18@gmail.com

Moscow, Lomonosov MSU, Faculty of Computational Mathematics and Cybernetics

Along with retina and fingerprints, using dorsal hand surface for identity recognition is one of the most promising directions of research in modern biometrics. Earlier the author presented a mobile identification system based on Android application. This article describes effective method of feature extraction and further matching running with EER = 0.5% on real data set (486 images of 42 people).

Feature generation is based on 2D Gabor filter whose form gives a possibility of analysis of the main finger knuckles characteristics: thickness, direction, mutual arrangement.

Various types of knuckles textural features have different characteristic sizes. Thus, firstly, two families of filters with different parameters σ (scaling) and λ (proportionally to the “period” of quasi-periodic structure of finger knuckles) were used. Assuming that, family is a set $\{G_R(x, y, \theta_j, \dots), j = 0, 1, \dots, 5\}$ where $G_R(\dots)$ — the real part of Gabor filter kernel with fixed parameters $\sigma, \gamma, \lambda; \theta_j$ — rotation angle, $\theta_j = j\pi/6$.

Secondly, feature vector was generated taking mesh nodes (mesh steps are proportional to corresponding sizes of filter window), not all the points from an image. Experimentally proved that such modification doesn't worsen identification quality but significantly reduces two features vectors comparison time.

- [1] *Chernyshov V.* Efficient Hand Detection on Client-server Recognition System // Proceedings of the 10th International Conference on Computer Vision Theory and Applications - Volume 2, Berlin: INSTICC, 2015. — p. 461–468. <http://www.scitepress.org/DigitalLibrary/PublicationsDetail.aspx?ID=PzA/cNUbpXE=&t=1>.

Исследование движения структуры радужной оболочки глаза методом оптического потока

Чигринский Виктор Владимирович^{1*}

chigrinskiy.viktor@phystech.edu

Матвеев Иван Алексеевич^{1,2}

matveev@ccas.ru

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Федеральный исследовательский центр «Информатика и управление» РАН

Распознавание по радужке — один из основных современных методов автоматической биометрической аутентификации. На изображении глаза или последовательности таких изображений (видео), выделяется область радужки, обладающая уникальным устойчивым рисунком. Перед работой алгоритмов сравнения происходит нормировка этой области — приведение её к стандартным размерам, что делается линейным относительно её радиуса преобразованием. Здесь возникает проблема, состоящая в том, что деформация ткани радужки не является линейной.

В работе исследуется нелинейное движение элементов радужной оболочки глаза при изменении размеров зрачка. Для слежения за перемещением элементов радужки используются методы оптического потока Лукаса–Канаде, Хорна–Шунка, а также их модификации. Целью работы является оценка вида радиально-симметричной функции, описывающей положение элементов радужной оболочки в зависимости от радиуса зрачка. Для оценки применимости метода оптического потока предлагается генерация синтетических данных в соответствии с заранее выбранной моделью деформации радужки, реализация метода на построенных данных и сравнение полученных результатов с ожидаемыми. Для тестирования алгоритма используется видео реакции человеческого глаза на вспышку света, полученное с помощью специального прибора.

Работа поддержана грантом РФФИ № 16-07-01171.

- [1] *Чигринский В. В., Матвеев И. А.* Исследование движения структуры радужной оболочки глаза методом оптического потока // Машинное обучение и анализ данных, 2017.

Iris structure motion analysis via optical flow method

Chigrinskiy Viktor^{1*}

chigrinskiy.viktor@phystech.edu

Matveev Ivan^{1,2}

matveev@ccas.ru

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

²Moscow, Federal Research Center “Computer Science and Control”
of RAS

Iris recognition is one of the modern methods of automatical human biometrical authentication. Iris region with unique stable structure is detected in eye image or video. In the preprocessing step, region normalization is implemented via linear transformation. The problem is that in fact iris deformation is non-linear.

Nonlinear movements of elements of human iris during pupil size variations is studied. Tracking of iris elements is done with the help of Lucas–Kanade, Horn–Schunck optical flow methods and their modifications. The aim is to estimate a radially symmetric function which describes positions of iris structural elements with respect to pupil size. The quality of the method is assessed by applying it to the synthetic data, which is built from preselected deformation model, after that obtained function is matched against the model. To test the algorithm on real data video of human’s eye reaction on flashlight is used, which was recorded by a pupillometer device.

This research is funded by RFBR, grant 16-07-01171.

- [1] *Chigrinskiy V., Matveev I.* Iris structure motion analysis via optical flow method // *Journal of Machine Learning and Data Analysis*, 2017.

Условные морфологические меры сложности и информативности изображений

Брянский Станислав Андреевич^{1,2*} sbrianskiy@gosniias.ru
*Визильтер Юрий Валентинович*² viz@gosniias.ru

¹Москва, Московский авиационный институт (НИУ)

²Москва, ФГУП «Государственный научно-исследовательский институт авиационных систем» (ФГУП «ГосНИИАС»)

Недостатком известных ранее мер сложности и информативности изображений является то, что они зависят от площадей областей разбиения кадра, но не от их формы. Чтобы преодолеть этот недостаток, предлагается объединить аппарат мозаичных моделей формы Пытьева с аппаратом карт толщин, рассматривавшихся в рамках ММ Серра. Картой толщин фигуры F относительно набора структурирующих элементов (СЭ) B называется функция $t_{F,B}(x, y)$, которая каждой точке фигуры ставит в соответствие площадь максимального покрывающего ее элемента из B , целиком вписанного в фигуру F . Картой толщин мозаичной формы $F = F_1, \dots, F_n$ называется функция $\tau_{F,B}(x, y)$, которая каждой точке кадра ставит в соответствие $\tau_{F_i,B}(x, y) : (x, y) \in F_i$. В работе показано, что введенная ранее (Визильтер, Рубис, 2014) мера сложности мозаичной формы $\mu_H(F)$ может интерпретироваться как средняя нормированная площадь максимального аморфного СЭ, целиком заполняющего область принадлежности точки. Исходя из этого, обобщенную условную морфологическую сложность $\mu_{H,B}(F)$ формы F относительно набора структурирующих элементов B на кадре площади S , можно определить как функцию $\mu_{H,B}(F) = 1 - \text{mean}_\omega(\tau_{F,B}(x, y))/S$. Условную морфологическую информативность $I_B(F)$ по аналогии можно определить как $I_B(F) = -\text{mean}_\omega(\log(\tau_{F,B}(x, y)/S))$. Подобным образом определяются взаимная морфологическая сложность и информативность двух изображений. В работе приведены примеры вычисления всех предложенных характеристик формы.

Работа выполнена при финансовой поддержке РФФ, грант № 16-11-00082.

[1] *Брянский С. А., Визильтер Ю. В.* Условные морфологические меры сложности и информативности изображений // ВКИТ., 2017 (в печати).

Morphology measures of complexity and information for binary images

Brianskiy Stanislav^{1*}
*Vizilter Yury*²

sbrianskiy@gosniias.ru
viz@gosniias.ru

¹Moscow, Moscow Aviation Institute (NRU)

²Moscow, The Federal State Unitary Enterprise “State Research Institute of Aviation Systems” (FGUP “GosNIIAS”)

Defect known measures of complexity and information for images is addition from squares of frame areas, but not of their shape. To overcome this defect, it is proposed combine mosaic models of Pyt'ev and thickness maps from mathematical morphology of Serra. Thickness map of shape F for set of structuring elements (SE) is the function $t_{F,B}(x,y)$, which associates each point of the shape and square of the maximal covering element B in the figure F . Thickness map of mosaic form $F = F_1, \dots, F_n$ is the function $\tau_{F,B}(x,y)$, which associates each point of the frame and function $\tau_{F_i,B}(x,y) : (x,y) \in F_i$. This paper is shown that the measure of complexity of mosaic form $\mu_H(F)$ introduced (Vizilter, Rubis, 2014) can be interpreted as the average normalize square of the maximum amorphous SE, which filling the domain of the point. Morphology measure of complexity $\mu_{H,B}(F)$ shape F for set of SE B on frame with square S can be defined as the function $\mu_{H,B}(F) = 1 - \text{mean}_\omega(\tau_{F,B}(x,y))/S$. Morphology measure of information $I_B(F)$ can be defined as $I_B(F) = -\text{mean}_\omega(\log(\tau_{F,B}(x,y)/S))$. Similary, the mutual measure of complexity and information for two images can be determined. In the work calculation all offered characteristics of the shape are given.

This research is funded by RNF, grant 16-11-00082.

- [1] *Brianskiy S. A., Vizilter Yu. V.* Morphology measures of complexity and information for binary images // Herald of computer and information technologies, 2017 (in press).

Поиск изображений по визуальным характеристикам с использованием гранично-скелетной модели формы

*Жукова Ксения Владимировна*¹

kz@pisem.net

*Рейер Иван Александрович*¹*

reyer@forecsys.ru

*Инякин Андрей Сергеевич*¹

inyakin@forecsys.ru

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

В докладе предлагается подход к поиску изображений в базах графических данных по визуальным характеристикам. Используется непрерывная модель сегментированного изображения, состоящая из набора непересекающихся многоугольных фигур. Каждый многоугольник из этого набора аппроксимирует выделенную растровую область, при этом многоугольники двух соседних областей имеют общие фрагменты границы. Для построения множества многоугольников используется модификация алгоритма аппроксимации бинарного изображения разделяющими многоугольниками минимального периметра. Сегментация изображения основана на процедурах выделения суперпикселей — небольших однородных областей. Соответствующие суперпикселям многоугольники объединяются в относительно небольшое количество крупных фигур с использованием алгоритма иерархической кластеризации цветов суперпикселей и анализа количества и соседства элементов в выделенных кластерах. Модель изображения также включает «размеченные» скелеты многоугольников, описывающие изменение скелетного представления, и оценки значимости выпуклых особенностей границы, соответствующих вершинам многоугольников [1]. Оценки значимости вычисляются на основе анализа параметрического семейства гранично-скелетных моделей формы, порожденного многоугольной фигурой. Полученные модели изображений сравниваются по форме и цвету многоугольных фигур. Для оценки сходства формы сравнивается характер изменения числа существенных особенностей фигур при росте величины точности аппроксимации.

Работа поддержана грантами РФФИ № 14-07-00736 и № 17-07-01432.

- [1] *Reyer I., Zhukova K.* Parametric Shape Descriptor based on a Scalable Boundary-Skeleton Model // *Communications in Computer and Information Science*, 2017 (в печати).

Region-based image retrieval with use of boundary-skeleton shape model

*Zhukova Ksenia*¹

kz@pise.net

*Reyer Ivan*¹*

reyer@forecsys.ru

*Inyakin Andrey*¹

inyakin@forecsys.ru

¹Moscow, Federal Research Center “Computer Science and Control” of RAS

An approach to region-based image retrieval is suggested. A continuous model of a segmented image consisting of a set of nonoverlapping polygonal figures is constructed. Each polygon from the set approximates a segmented raster region within the image, with polygons of two neighbor regions having common fragments of boundary. To obtain the set of polygons a modified algorithm for approximation of a binary image with polygons of minimal perimeter is used. The segmentation is based on superpixel extraction procedures. The polygons corresponding to superpixels are grouped into a relatively small number of large figures using an algorithm of hierarchical color clustering and analyzing the number and neighborhood of elements in the obtained clusters. The model also includes marked skeletons of polygons describing changes of skeletal representation and significance estimations for boundary convexities corresponding to polygon vertices [1]. The estimations are calculated with use of a family of boundary-skeleton shape models generated by a polygonal figure. Obtained image models are compared by shape and color of polygons. To estimate the shape similarity, the change of significant convexities' number at increase of the approximation accuracy value is compared.

This research is funded by RFBR, grants 14-07-00736 and 17-07-01432.

- [1] *Reyer I., Zhukova K.* Parametric Shape Descriptor based on a Scalable Boundary-Skeleton Model // Communications in Computer and Information Science, 2017 (in press).

Распараллеливание алгоритма сравнения бинарных изображений, представленных цепочками скелетных примитивов

*Кушнир Олеся Александровна*¹ kushnir-olesya@rambler.ru
Середин Олег Сергеевич^{1*} oseredin@yandex.ru

¹Тула, Тульский государственный университет

В тех задачах компьютерного зрения, где важным дискриминантным признаком объектов является форма, для описания объекта может использоваться его скелет. Сопоставление между собой двух скелетов с целью определения меры сходства или различия объектов является ресурсоёмкой в вычислительном плане задачей. Зачастую, необходимость в сопоставлении объектов возникает не только при обработке отдельных изображений, но и при обработке видеопотока, а это накладывает дополнительные ограничения на скорость работы алгоритма сравнения объектов на каждом кадре видеоряда. Добиться ускорения работы алгоритма можно двумя способами: аналитически, улучшая алгоритм и сокращая количество выполняемых им операций; и технически, путём параллельного выполнения некоторых участков алгоритма.

В данной работе будет представлено решение задачи ускорения сравнения скелетов с точки зрения возможности ее распараллеливания. Будут рассмотрены способы ускорения работы ранее предложенного метода сравнения бинарных изображений на основе описания скелетов объектов цепочками примитивов; обосновано применение параллельных вычислений на персональном компьютере с использованием технологии OpenMP и библиотек для работы с потоками стандарта языка программирования C++11. Приведены результаты экспериментов по оценке ускорения, полученного параллельными реализациями алгоритма сравнения бинарных изображений, относительно последовательной версии [1].

Работа поддержана грантом РФФИ № 16-57-52042.

- [1] *Кушнир О. А., Середин О. С.* Параллельные реализации алгоритма сравнения бинарных изображений на основе описания скелетов цепочками примитивов // Вестник компьютерных и информационных технологий, 2017 (в печати).

The Parallel Implementations of the Shape Comparison Algorithm Based on Skeleton Primitive Chains Alignment

*Kushnir Olesia*¹

kushnir-olesya@rambler.ru

Seredin Oleg^{1*}

oseredin@yandex.ru

¹Tula, Tula State University

The skeleton is a useful descriptor for solving Computer Vision problems based on shape matching. Skeleton matching is a rather computationally expensive task, so some ways of reducing computations are appreciated, especially in case of video stream processing. There are two main methods of speeding up an algorithm – the first one is the analytical, by improving the algorithm and reducing the number of operations in it; and the second one is the technical, by parallelizing some parts of the source code.

In this work we propose the technical way of speeding up the previously suggested skeleton matching algorithm based on skeleton primitive chains alignment. We review two parallel computing techniques for the systems with shared memory – OpenMP and C++11 multi-threading – and prove that they are proper for our skeleton matching algorithm which could be used on ordinary multi-core PCs. Then we show which part of the algorithm would be parallelized. As a conclusion we discuss experimental results achieved on two parallelized versions and a sequential version of the algorithm [1].

This research is funded by RFBR, grant 16-57-52042.

- [1] *Kushnir O., Seredin O.* The Parallel Implementations of the Binary Image Comparison Algorithm Based on Skeleton Primitive Chains // *Jornal of Computer and Information Technology*, 2017 (in press).

Выделение отличий на изображениях с помощью референтных EMD-фильтров

<i>Лебедев Максим Алексеевич</i> ^{1*}	MLebedev@gosniias.ru
<i>Рубис Алексей Юрьевич</i> ¹	arubis@gosniias.ru
<i>Визильтер Юрий Валентинович</i> ¹	viz@gosniias.ru
<i>Выголов Олег Вячеславович</i> ¹	o.vygolov@gosniias.ru
<i>Желтов Сергей Юрьевич</i> ¹	zhl@gosniias.ru

¹Москва, ФГУП «Государственный научно-исследовательский институт авиационных систем» (ФГУП «ГосНИИАС»)

Ранее для задачи выделения отличий в случае существенно-го изменения как яркости, так и геометрии наблюдаемой сцены, была предложена морфологическая схема анализа изображения, основанная на диффузных и референтных мозаичных фильтрах, определяемых тепловыми ядрами сходства областей разбиения кадра. В данной работе для реализации этой схемы предлагаются оригинальные референтные EMD-фильтры, основанные на определении оптимальных матриц взаимного сходства мозаичных форм методом линейного программирования.

Предложенная схема референтной EMD-фильтрации была реализована и протестирована применительно к ряду модельных и реальных изображений. Эксперименты показали существенное улучшение качества выделения отличий по сравнению с реализованной ранее схемой референтной фильтрации на основе диффузных референтных фильтров с матрицами сходства, определяемыми тепловыми ядрами. Существенное преимущество использования EMD-фильтрации в морфологических схемах сравнения изображений заключается в том, что взаимные и референтные фильтры с EMD-матрицей взаимного сходства являются проекторами (идемпотентными операторами). Таким образом, EMD-морфология принимает в точности такой же вид, как классическая морфология Пытьева.

Работа поддержана грантом РНФ № 16-11-00082.

- [1] *Лебедев М. А. Рубис А. Ю. Визильтер Ю. В. Выголов О. В. Желтов С. Ю.* Выделение отличий на изображениях с помощью референтных EMD-фильтров // Вестник компьютерных и информационных технологий, Москва: ООО «Издательский дом «Спектр», 2017 (в печати).

Image differences detection based on reference EMD-filters

Lebedev Maksim^{1*}

MLebedev@gosniias.ru

*Rubis Aleksey*¹

arcelt@mail.ru

*Vizilter Yuri*¹

viz@gosniias.ru

*Vygolov Oleg*¹

o.vygolov@gosniias.ru

*Zheltoy Sergey*¹

zhl@gosniias.ru

¹Moscow, The Federal State Unitary Enterprise “State Research Institute of Aviation Systems” (FGUP “GosNIIAS”)

In our previous works, dedicated to differences detection on images in case of substantial brightness and object geometry variation, we proposed a methodology of morphological image analysis based on diffusion and reference mosaic filters, which were defined through the heat kernels of image areas similarity. In this paper, we propose an implementation of this methodology using original reference EMD-filters based on calculation of optimal matrices of mosaic forms mutual similarity through linear programming method.

The proposed algorithm of reference EMD-filtration was tested on a set of real and synthetic images. The experiments have shown the significant increase in quality of image differences detection compared with the early-developed algorithm of reference filtering based on diffusion reference filters with similarity matrices and heat kernels. The substantial advantage of embodying EMD-filters in morphological image comparison schemes is due to the mutual and reference filters with EMD-matrix of mutual similarity being the projectors (idempotent operators). As a result, EMD-morphology exactly corresponds to classical Pyt'ev morphology.

This research is funded by RSF, grant 16-11-00082.

- [1] *Lebedev M. Rubis A. Vizilter Yu. Vygolov O. Zheltoy S.* Image differences detection based on reference EMD-filters // Herald of computer and information technologies, Moscow: Publishing house “Spektr”, 2017 (in press)

Классификация бинарных фигур с использованием мешка контекстов ребер скелета

Ломов Никита Александрович^{1,2}

nikita-lomov@mail.ru

Сидякин Сергей Владимирович^{2*}

sersid@gosniias.ru

¹Москва, Московский государственный университет

им. М. В. Ломоносова

²Москва, ФГУП «ГосНИИАС»

В статье рассматривается задача классификации бинарных фигур. Для ее решения широко используются контурные и скелетные подходы. При этом полностью контурные методы более успешны чем полностью скелетные. На сегодняшний день наибольший прогресс достигнут в случае явного или косвенного обогащения контурных представлений скелетной информацией. В данной работе предлагается два новых полностью скелетных дескриптора бинарной фигуры – контекст ребра скелета и мешок контекстов ребер скелета. Контекст ребра скелета является обобщением контекста формы на скелет и содержит информацию обо всех особенностях формы фигуры с «локальной» точки зрения. Если гибкие деформации не меняют топологических характеристик фигуры, а локальные толщины не изменяются слишком сильно, то предложенный дескриптор оказывается устойчивым к нежестким искажениям частей фигуры. Контексты ребер кодируется на основе мешка признаков в результате чего формируется мешок контекстов ребер скелета. Мешок контекстов ребер содержит глобальную информацию обо всех особенностях формы фигуры со всех «локальных» точек зрения. Комбинация мешка контекстов ребер и контурного дескриптора существенно улучшает качество классификации на сложных коллекциях изображений. В работе также описаны эффективные алгоритмы построения предложенных дескрипторов на основе непрерывных скелетных моделей.

Работа поддержана грантом РФФ № 16-11-00082 и грантами РФФИ № 15-07-01323 А и № 16-57-52042 МНТ_а.

- [1] *Lomov N., Sidiyakin S.* Shape classification using bag of edge contexts // Computer Vision and Image Understanding, USA: Elsevier Inc., 2017.

Shape classification using bag of edge contexts

Lomov Nikita^{1,2}

nikita-lomov@mail.ru

Sidyakin Sergey^{2*}

sersid@gosniias.ru

¹Moscow, Lomonosov Moscow State University

²Moscow, FGUP “GosNIIAS”

This paper addresses 2D shape representation and classification issues. Generally, the existing main shape representations are contour-based and skeleton-based. Pure contour-based representations are more successful than pure skeleton-based ones. However, the current progress in shape classification is bound to contour representations enriched with the information from skeleton either implicitly or explicitly. In this paper we propose a novel skeleton-based shape descriptor named Edge Context. Edge contexts retain information about the global structure of the entire shape from a local point of view. If non-rigid deformations do not break the topology and local thicknesses are not changed too much they are stable to flexible deformations of the shape parts. Edge contexts are encoded in the Bag of Features framework and Bag of Edge Contexts is produced. Such descriptor can explicitly improve a contour descriptor when both of them are combined into a single feature vector. We also provide quite efficient algorithms for construction of the mentioned descriptors using a continuous skeleton. Experimentation results indicate that proposed shape representation achieves the state-of-the-art performances on several challenging shape datasets.

This work is supported by Russian Science Foundation (RSF) under Grant 16-11-00082; by Russian Foundation For Basic Research (RFBR) under Grants 15-07-01323 A and 16- 57-52042 MHT a.

[1] *Lomov N., Sidyakin S.* Shape classification using bag of edge contexts // Computer Vision and Image Understanding, USA: Elsevier Inc., 2017.

Морфологическая оценка динамики регенерации планарий по фотоизображениям

Местецкий Леонид Моисеевич^{1*}
*Тирас Харлампий Пантелеевич*²

author_mestlm@mail.ru
tiras1950@yandex.ru

¹Москва, МГУ им. М. В. Ломоносова, Факультет ВМК

²Пушино, Пушкинский естественно-научный институт

Пресноводные ресничные плоские черви планарии обладают уникальной способностью к регенерации – отрастанию части тела, утраченной в результате поперечной перерезки. Скорость регенерации зависит от внешних факторов, поэтому планарии используются в качестве экспериментальных животных в экологических и биологических исследованиях. Для таких исследований необходим объективный количественный контроль скорости регенерации. Сложность контроля определяется тем, что планарии не обладают жестким наружным или внутренним скелетом, они двигаются и быстро изменяют свою формы в ходе движения [1]. В докладе рассматривается задача оценивания скорости регенерации на основе последовательности фотографий, полученных в течение нескольких дней наблюдения за группой планарий. Предлагается подход, основанный на построении непрерывной морфологической модели силуэта планарии и анализе дискового покрытия силуэта – объединении всех вписанных в фигуру дисков заданного размера [2]. Предложен количественный критерий, основанный на вычислении площади дискового покрытия. Динамика морфогенеза оценивается по изменению площади покрытия за определенные интервалы времени. Разработан численный метод, реализующий этот подход. Корректность и эффективность метода подтверждается результатами вычислительных экспериментов на большой базе изображений. Работа поддержана грантами РФФИ №№ 17-01-00917, 15-29-01135.

- [1] *Тирас Х. П. и др* Минимизация погрешностей морфометрии регенерирующих планарий // *Фундаментальные исследования*, No. 2, ч.7, 2015. — С. 1412-1416.
- [2] *Ломов Н. А., Местецкий Л. М.* Площадь дискового покрытия – дескриптор формы изображения // *Компьютерная оптика*, 40(4), 2016. — С. 516–525.

Morphological evaluation of the dynamics of planarian regeneration through photo-images

Mestetskiy Leonid^{1*}

mestlm@mail.ru

*Tiras Kharlampiy*²

tiras1950@yandex.ru

¹Moscow, Lomonosov MSU, Faculty of Computational Mathematics and Cybernetics

²Pushchino, Pushchino State Institute of Natural Sciences

Freshwater ciliated planar worms have a unique ability to regenerate the part of the body that is lost as a result of transverse transection. The speed of regeneration depends on external factors, therefore planarians are used as experimental animals in ecological and biological studies. For such studies, objective quantitative control of the rate of regeneration is necessary. The complexity of control is determined by the fact that the planarians do not have a rigid outer or inner skeleton, they move and rapidly change their shape during the movement [1].

In the report, the problem of estimating the rate of regeneration is considered on the basis of a sequence of photographs obtained during a few days of observation of a group of planarians. An approach is proposed for estimating the rate of regeneration from the available digital gray-scale images, based on the construction of a continuous morphological model of the planarian silhouette and the analysis of the disk cover of the silhouette (the union of all disks of a given size inscribed in the figure) [2]. The process of regeneration is expressed as the transformation of the semicircular shape of the growing part of the planar body into a triangular shape. The proposed quantitative criterion is based on the calculation of the disc cover area. The dynamics of morphogenesis is estimated by changing the area during certain time intervals. A numerical method has been developed that implements this approach. The correctness and efficiency of the method is confirmed by the results of computational experiments on a large image base. This work is supported by RFBR, grants 17-01-00917, 15-29-01135.

[1] *Tiras K. P. et al.* Minimization of errors in the morphometry of regenerating planarians // Basic research, 2(7), 2015. — P. 1412–1416.

[2] *Lomov N. A., Mestetskiy L. M.* Area of the disk cover as an image shape descriptor // Computer Optics, 40(4), 2016. — P. 516–525.

Метод выделения подмножества векторных представлений утолщенных горизонталей на растровых топографических картах

*Мирошниченко Сергей Юрьевич*¹ oldguy7@rambler.ru

Дремов Евгений Николаевич^{1*} evgeni-dremov@yandex.ru

¹Курск, Юго-Западный государственный университет

Разработан метод автоматического выделения подмножества утолщенных горизонталей из множества векторных представлений изолиний, сформированного по растровым топографическим картам. Метод основан на яркостной оценке ширины линии и ее сравнении с оценками смежных изолиний. Рассмотрены основные этапы работы предлагаемого метода, приведены результаты его применения для автоматического выделения утолщенных горизонталей на топографических картах.

Предлагаемый метод относится к области картографии и применяется при создании и актуализации цифровых моделей рельефа с использованием в качестве исходного материала растровых топографических карт.

Точность предложенного метода подтверждена в результате экспериментальных исследований на сериях топографических карт. Результаты экспериментальных исследований подтверждают целесообразность использования предложенного метода для формирования множества векторных представлений утолщенных горизонталей на растровых топографических картах.

Таким образом, представленный метод позволяет посредством анализа топографической карты выделять утолщенные горизонталю из общего множества векторных представлений изолиний и может быть использован при автоматическом создании цифровых моделей рельефа.

- [1] *Мирошниченко С. Ю.* Метод выделения подмножества векторных представлений утолщенных горизонталей на растровых топографических картах // Телекоммуникации, Москва: Наука и технологии, 2017. — С. 37–42.

Method selection subset of the vector representation thickening horizontals on raster topographic maps

*Miroshnichenko Sergey*¹

oldguy7@rambler.ru

Dremov Evgeny^{1*}

evgeni-dremov@yandex.ru

¹Kursk, Southwest State University

A method has been developed to automatically selection a subset of thickened contours from a set of vector representations of isolines formed by raster topographic maps. The method is based on the brightness-based evaluation of isoline width and its comparison to adjacent lines. The main steps of the proposed method, the results of its application for the automatic thickened isolines detection on topographic maps are shown and described.

The proposed method refers to the field of cartography and is used in the creation and actualization of digital relief models using raster topographic maps as a raw material.

The accuracy of the proposed method is confirmed by experimental studies on a series of topographic maps. The results of experimental studies confirm the expediency of using the proposed method for forming a set of vector representations of thickened contour lines on raster topographic maps.

Thus, the presented method allows, through the analysis of a topographic map, to isolate thickened horizons from a common set of vector isolines and can be used in the automatic creation of digital relief models.

- [1] *Miroshnichenko S. Yu.* Method selection subset of the vector representation thickening horizontals on raster topographic maps // Telecommunications, Moscow: Science and Technology, 2017. — p. 37–42.

Морфологический проектор в метрике L_0 и задача поиска структурных различий изображений

*Перевалов Денис Сергеевич*¹

perevalovds@gmail.com

Корнилов Фёдор Андреевич^{1*}

fakornilov@mail.ru

¹Екатеринбург, Институт математики и механики
им. Н. Н. Красовского УрО РАН

Работа посвящена задаче обнаружения и локализации структурных различий изображений, под которыми понимаются изменения в составе объектов сцены. Для случая конечнозначных изображений предложен новый алгоритм вычисления области различий, основанный на морфологической проекции в метрике L_0 , и показано, что он дает точное решение для достаточно широкого класса структурных различий. Получено утверждение о том, что для изображений, зашумленных дискретным независимым нормальным белым шумом, симметрическая мера разности результата работы предложенного алгоритма и истинного множества различий стремится по вероятности к нулю при неограниченном росте отношения величины минимального скачка яркости к СКО шума. На основе данного алгоритма разработан его регуляризованный вариант, опирающийся на идею построения ряда проекций для различных множеств уровня, комплексно нивелирующее влияние шума, и их последующего объединения в результирующее изображение. Приводятся результаты сравнения качества работы предложенных алгоритмов и известных методов поиска различий на реальных и синтезированных изображениях, демонстрирующие применимость предложенного подхода для решения прикладных задач.

Работа выполнена при частичной поддержке комплексной программы ФНИ УрО РАН (проект 15-16-1-14).

- [1] *Костоусов В. В., Перевалов Д. С.* Морфологический проектор в метрике L_0 и задача локализации структурных различий изображений // Труды Института математики и механики УрО РАН, Екатеринбург: ИММ УрО РАН, Том 23, № 1, 2017. — С. 171–187.

Morphological projector in the L_0 metric and the problem of the finding structural differences between images

*Perevalov Denis*¹

perevalovds@gmail.com

Kornilov Fedor^{1*}

fakornilov@mail.ru

¹Ekaterinburg, Krasovskii Institute of Mathematics and Mechanics
UB RAS

The paper is devoted to the problem of the detection and localization of the structural differences between images, by which are meant changes in the composition of scene objects. For the case of the finite-valued images the new algorithm for estimating the area of the differences are proposed. The algorithm is based on the morphological projection in the L_0 metric. It is shown, that the algorithm gives the exact solution for a fairly wide class of the structural differences. Also it is proven that for images noised by discrete independent normal white noise, symmetric measure of the difference in the result of the proposed algorithm and a true differences area goes to the zero in probability, in the case, when the ratio of the minimum brightness jump to the SNR of the noise infinitely grows. Based on this algorithm, it was developed its regularized variant, which utilizes the idea of constructing a series of projections for various level sets, which complexly decrease the noise effect. This level sets subsequently integrated into the resulting image, which is output of the regularized algorithm. The results of a comparison of the quality of work of the proposed algorithms and known methods for finding differences on real and synthesized images are presented. They demonstrate the applicability of the proposed approach for solving applied problems.

The work was carried out with the partial support of a comprehensive program of the Fundamental Scientific Researches by the UB RAS (project 15-16-1-14).

- [1] *Kostousov V., Perevalov D.* Morphological projector in the L_0 metric and the problem of localization of structural differences between images // Proc. of the Krasovskii Institute of Mathematics and Mechanics, Ekaterinburg, Vol. 23, Is. 1, 2017. — pp. 171–187.

Компаративная фильтрация изображений с использованием монотонных морфологических операторов

Рубис Алексей Юрьевич^{1*}

arubis@gosniias.ru

*Лебедев Максим Алексеевич*¹

MLebedev@gosniias.ru

*Визильтер Юрий Валентинович*¹

viz@gosniias.ru

¹Москва, ФГУП «Государственный научно-исследовательский институт авиационных систем» (ФГУП «ГосНИИАС»)

Ранее в работе (Vizilter et al., 2016) была предложена схема построения компаративных морфологий, обобщающая морфологический подход Ю.П. Пытьева на более широкий класс процедур, решающих практические задачи сравнения изображений. В рамках этой схемы были разработаны процедура фильтрации на основе взаимного контрастирования и алгоритм выделения отличий на ее основе.

В данной работе представлен класс компаративных фильтров, использующих в качестве упрощающих фильтров операции математической морфологии (ММ) Серра, которые, в отличие от рассматривавшихся нами ранее проекторов Пытьева, линейных диффузных и нелинейных медианных фильтров, являются монотонно неувеличивающими (неуменьшающими) фильтрами:

$$C(f, g) = S(f) + a(f, g) \cdot [f - S(f)]_t, a(f, g) \in \{0, 1\},$$

где $[f - S(f)]_t$ – отсечение значения разности по минимальному порогу t ; f, g – тестовое и эталонное изображения соответственно; $a(f, g)$ – локальный коэффициент взаимного сходства изображений f и g ; $S(f)$ – морфологический фильтр Серра.

Приведены результаты экспериментального исследования предложенных фильтров с различными параметрами в задаче выделения отличий на изображениях в сравнении другими морфологическими процедурами выделения отличий.

Работа поддержана грантом РФФИ № 16-11-00082.

- [1] Рубис А. Ю. Лебедев М. А. Визильтер Ю. В. Компаративная фильтрация изображений с использованием монотонных морфологических операторов // Вестник компьютерных и информационных технологий, Москва: ООО «Издательский дом «Спектр», 2017 (в печати).

Comparative image filtering using monotonical morphology operators

Rubis Aleksey^{1*}

arubis@gosniias.ru

*Lebedev Maksim*¹

MLebedev@gosniias.ru

*Vizilter Yuri*¹

viz@gosniias.ru

¹Moscow, The Federal State Unitary Enterprise “State Research Institute of Aviation Systems” (FGUP “GosNIIAS”)

In previous work (Vizilter et al., 2016) Comparative Morphology (CM) construction scheme generalized Pyt’ev’s Morphological Image Analysis approach to a wider area of practical image comparison applications was proposed. In the framework Guided Contrasting filtration procedure and change detection algorithm based on it were developed.

In this work we propose a class of CM filters using operators of Serra’s Mathematical Morphology as smoothing operators, that are monotonically non-increasing (non-decreasing) filters in contrast to linear diffuse filters and non-linear median filters:

$$C(f, g) = S(f) + a(f, g) \cdot [f - S(f)]_t, a(f, g) \in \{0, 1\},$$

where $[f - S(f)]_t$ – difference value after minimum thresholding by t ; f, g – reference and testing images respectively; $a(f, g)$ – local similarity measure of f and g images; $S(f)$ – Serra’s morphological filter.

Therefore, in the work results of change detection experiments with new CM filters in comparison with other morphological procedures are represented.

The work is founded by RSF Grant 16-11-00082.

- [1] *Rubis A. Yu. Lebedev M. A. Vizilter Yu. V.* Comparative image filtering using monotonical morphology operators (in russian) // Herald of computer and information technologies, Moscow: Publishing house “Spektr” , 2017 (in press).

Детектирование переводных заимствований в текстах научных статей из журналов, входящих в РИНЦ

Бахтеев Олег Юрьевич^{1,2*}

bakhteev@ap-team.ru

Кузнецова Маргарита Валерьевна^{1,2}

kuznetsova@ap-team.ru

¹Москва, ЗАО «Антиплагиат»

²Москва, Московский физико-технический институт (ГУ)

Рассматривается задача детектирования переводных заимствований. Такой вид заимствований, когда изготовитель работы, с использованием средств машинного перевода или самостоятельно переводит преимущественно англоязычные научные публикации на русский язык и выдает их за оригинальные, становится наиболее популярным из-за нехватки промышленных инструментов, детектирующих заимствования такого рода.

Авторами предложен алгоритм детектирования переводных заимствований для языковой пары «английский-русский». Алгоритм объединяет в себе отбор документов-кандидатов на заимствования и последующее их сравнение с проверяемым документом. В связи со спецификой рассматриваемой задачи, алгоритм должен быть устойчив к неоднозначностям перевода. В отборе документов-кандидатов эта устойчивость достигается за счет замены слов на метки классов эквивалентности, полученных с помощью дистрибутивной модели. Далее, для сравнения найденных кандидатов и рассматриваемого документа, используется отображение текстовых фрагментов документов в векторное пространство высокой размерности. Для этого отображения используются рекуррентные сети глубокого обучения. Для оптимизации параметров сети применяются технологии обучения без учителя и технологии частичного обучения.

В докладе представлены результаты работы предлагаемого алгоритма на статьях из журналов, входящих в Российский индекс научного цитирования (РИНЦ).

- [1] *Bakhteev O., Kuznetsova R., Romanov A., Khritankov A., A monolingual approach to detection of text reuse in Russian-English collection // Artificial Intelligence and Natural Language and Information Extraction, Social Media and Web Search FRUCT Conference (AINL-ISMW FRUCT), 2015. — С. 3–10.*

Cross-lingual plagiarism detection in the collection of scientific articles from journals included in RSI

Bakhteev Oleg^{1,2}

bakhteev@ap-team.ru

Kuznetsova Rita^{1,2}

kuznetsova@ap-team.ru

¹Moscow, JSC Antiplagiat

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

The paper describes the method of cross-lingual plagiarism detection. This kind of plagiarism became very popular because there are not tools for detections such kind of plagiarism.

The authors propose the algorithm of cross-lingual plagiarism detection for the language pair “english-russian”. Algorithm includes the source retrieval stage and following comparison between retrieved documents and the suspicious document. The algorithm should be stable to the translation ambiguity. At the source retrieval stage this ambiguity is achieved by moving from word shingles to word-class shingles, where each word is substituted by the label of the class obtained by distributive hypothesis. For comparison stage we use model for mapping the word sequence into high dimensional space. For this purpose we use the recurrent neural network. We train our model in unsupervised or semi-supervised training regimes.

We demonstrate the performance of the proposed algorithm on the collection of scientific articles from the journals included in RSI Index.

- [1] *Bakhteev O., Kuznetsova R., Romanov A., Khritankov A.*, A monolingual approach to detection of text reuse in Russian-English collection // Artificial Intelligence and Natural Language and Information Extraction, Social Media and Web Search FRUCT Conference (AINL-ISMW FRUCT), 2015. — p. 3–10.

Аддитивная регуляризация тематических моделей связного текста

Воронцов Константин Вячеславович voron@forecsys.ru
Москва, Московский физико-технический институт (ГУ)

Вероятностные тематические модели выявляют тематическую кластерную структуру текстовых коллекций. Модель представляет каждую тему распределением вероятностей слов и описывает каждый документ вероятностной смесью тем. Простейшие модели опираются на гипотезу «мешка слов», но известны также и модели, учитывающие порядок слов в тексте.

Модели n -грамм основаны на выделении информативных словосочетаний. Для этого используются статистические критерии коллокаций слов, автоматический синтаксический разбор, оценки тематичности n -грамм. Темы, образуемые n -граммами, намного лучше интерпретируются, чем темы в униграммных моделях.

Модели дистрибутивной семантики строят векторные представления слов по данным об их совместной встречаемости. В отличие от известных моделей word2vec и GloVe, тематические векторные представления разрежены и интерпретируемы.

Модели предложений и модели тематической сегментации основаны на гипотезе, что связный текст естественного языка состоит из последовательности сообщений, каждое из которых относится к одной теме или небольшому числу тем.

В докладе рассматриваются методы регуляризации тематических моделей перечисленных типов. Предлагается регуляризатор пост-обработки E-шага в EM-алгоритме для формализации широкого класса предположений о тематической структуре связного текста. Техника *аддитивной регуляризации* позволяет встраивать регуляризацию E-шага в тематические модели различных типов: иерархические, темпоральные, мультимодальные и др.

Рассматриваются приложения тематического моделирования для анализа записей разговоров контактного центра с использованием открытой библиотеки BigARTM, //http:bigartm.org.

Работа поддержана грантами РФФИ 17-07-01536, 16-37-00498.

- [1] Воронцов К. В. Обзор вероятностных тематических моделей // Автоматическая обработка текстов на естественном языке и анализ данных. — М.: Изд-во НИУ ВШЭ, 2017. — 272 с.

Additive regularization for topic modeling of coherent text

Vorontsov Konstantin

voron@forecsys.ru

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

Topic modeling is an area of text mining actively developed in last 15 years. A probabilistic topic model extracts a set of hidden topics from a collection of text documents. It represents each topic by a probability distribution over words and describes each document with a probabilistic mixture of topics.

The bag-of-words hypothesis is probably one of most criticized assumptions in basic topic models. In response to this criticism, many sequential extensions for modeling coherent text appeared.

The first one is based on considering *n-grams* or *collocations* of words rather than individual words. Topics inferred from *n-grams* are usually much easier to interpret than that based on unigrams. The problem of this approach is that the number of distinct *n-grams* grows very fast with the size of text collection.

The second extension is considering *co-occurrence* of words according to the Harris' distributional hypothesis. In contrast to GloVe, word2vec and other well-known *word embedding* models, the topic-based vector representations of words are sparse and interpretable.

The third extension is based on the assumption that the natural language text is usually a sequence of segments each containing only one topic or a very small number of topics. This leads to *sentence* topic models and topic-based *text segmentation* techniques.

We consider the above topic models within *additive regularization* framework. We propose the E-step regularization, which looks like E-step post-processing in the EM-algorithm to formalize assumptions about the topical structure of the coherent text. Additive regularization enables us to combine the E-step regularization with other types of topic models, e.g. hierarchical, temporal, multimodal, etc.

We report experimental results on the topic-based scenario analysis of call center records using the open-source library BigARTM available at [//http://bigartm.org](http://bigartm.org).

The work is supported by RFBR grants 17-07-01536, 16-37-00498.

- [1] *Kochedykov D., Apishev M., Golitsyn L., Vorontsov K.* Fast and modular regularized topic modeling // ISMW, 2017.

Формирование базы данных для системы оценки эмоционального фона диалога с оператором центра обработки вызовов

*Копылов Андрей Валериевич*¹

And.Kopylov@gmail.com

Середин Олег Сергеевич^{1*}

oseredin@yandex.ru

*Найденов Алексей Владимирович*²

anaidyonov@itoolabs.com

*Земин Дмитрий Геннадьевич*²

dzendmitry@itoolabs.com

¹Тула, Тульский государственный университет

²Тула, ITooLabs

Ключевым моментом для построения систем распознавания эмоций в речи является наличие размеченных баз данных, которые возможно использовать на этапе обучения. Для некоторых языков такие базы существуют, например Danish Emotional Speech corpus, Berlin Emotional Database, Spanish Emotional Speech Database, Chinese Emotion Speech Databases, Japanese Emotional Speech Database и др. Для русского языка единственная доступная база RUSLANA содержит записи, эмоциональное состояние которых смоделировано студентами-лингвистами в начале 2000-х, и основные классификаторы, как на основе традиционных методов (опорных векторов, скрытых марковских моделей), так и на основе методов глубокого обучения, показывают на данной базе крайне низкие результаты. Кроме того, способы представления эмоциональных состояний на основе пяти-семи базовых эмоций, наиболее часто используемые при формировании подобных баз, не слишком подходят для задач анализа диалогов в центрах обработки вызовов (call center).

В данной работе предлагается шкала эмоциональных состояний, характеризующая положительное или отрицательное отношение собеседника к разговору и выраженность недовольства в случае, если оно присутствует, а также методика разметки записей телефонных разговоров для создания базы данных эмоциональных состояний собеседников в ходе диалога [1].

- [1] *Копылов А. В., Середин О. С., Найденов А. В., Земин Д. Г.* Формирование базы данных для системы оценки эмоционального фона диалога с оператором центра обработки вызовов // Известия ТулГУ, Технические науки Тула: Издательство ТулГУ, 2017 (в печати).

The creation of a corpus of emotional data for the system of emotion-related states assessment of a dialogue with the call center operator

*Kopylov Andrei*¹

And.Kopylov@gmail.com

Seredin Oleg^{1*}

oseredin@yandex.ru

*Naidyonov Alexey*²

anaidyonov@itoolabs.com

*Zenin Dmitry*²

dzendmitry@itoolabs.com

¹Tula, Tula State University

²Tula, ITooLabs

The main objective for the development of systems of emotion recognition in speech is the presence of labeled emotional corpus that can be used at the training stage. For some languages, such databases exist, for example, Danish Emotional Speech Corpus, the Berlin Emotional Database, the Spanish Emotional Speech Database, the Chinese Emotion Speech Database, the Japanese Emotional Speech Database, etc. For the Russian language, the only available database RUSLANA contains entries, whose emotional state has been simulated by linguistics students in the early 2000s, and the main classifiers, both on the basis of traditional methods (SVM, Hidden Markov Models, etc.), and on the basis of deep learning methods, show extremely low results on this corpus. In addition, the ways of representing emotional states based on five–seven universal basic emotions, most often used in the formation of such emotional corpora, are not very suitable for analysis of dialogues in call centers.

In this paper we propose a scale of emotional states that characterizes the positive or negative attitude of the speaker and the activation level of negative emotion in case it is present, as well as the technique for labeling a record of telephone conversations to create a corpus of emotion-related states of speakers in the dialogue [1].

- [1] *Kopylov A. V., Seredin O. S., Naidyonov A. V., Zenin D. G.* The creation of a corpus of emotional data for the system of emotion-related states assessment of a dialogue with the call center operator // Proceedings of TulSU, Tula: Tula State University, 2017 (in press). .

Исследование и реализация рекомендательных веб-сервисов выбора образовательных ресурсов для реализации индивидуальных образовательных траекторий, обучающихся на основе коллаборативной фильтрации

Липко Юлия Юрьевна^{1*}

ylipko@sfnedu.ru

*Свиридов Александр Славьевич*¹

asviridov@sfnedu.ru

¹Таганрог, Южный федеральный университет

В работе осуществлено исследование методов моделирования систем выбора альтернатив по информации, опубликованной на образовательных веб-сервисах для реализации индивидуальных образовательных траекторий обучающихся на основе предпочтений освоивших предлагаемые образовательные ресурсы.

Актуальность исследования обусловлена необходимостью развития информационных технологий экспертизы и сортировки контента, в которой участвуют как специалисты в исследуемой предметной области, так и пользователи, при этом альтернатива выбирается при помощи коллективной фильтрации. Основой рассматриваемого метода коллаборативной фильтрации является предположение, что похожим пользователям веб-сервиса будут полезны похожие образовательные ресурсы. Входной информацией для данного метода является матрица студент – образовательный ресурс, в которой отражены пользовательские оценки, все оценки одного пользователя представляют собой вектор. Косинусная мера позволяет оценивать схожесть векторов по значению косинуса угла между ними.

Модель применена для совершенствования технологий самостоятельной работы студентов для реализации личного потенциала обучающегося при реализации веб-порталов образовательных ресурсов и организации дистанционного образования.

Работа поддержана грантом РФФИ № 15-07-04102.

- [1] *Липко Ю. Ю.* Исследование методов моделирования рекомендательных веб-сервисов выбора образовательных ресурсов для реализации индивидуальных образовательных траекторий, обучающихся на основе коллаборативной фильтрации // Известия ЮФУ. Технические науки, Ростов-на-Дону: Издательство ЮФУ, 2016. — С. 115–122.

Research and implementation of recommended web services for the selection of educational resources for the implementation of individual educational trajectories of students on the basis of collaborative filtering

Lipko Yulia^{1*}

ylipko@sfedu.ru

*Sviridov Alexander*¹

asviridov@sfedu.ru

¹Taganrog, Southern Federal University

In operation the research of methods of simulation of systems of a choice of alternatives on the information published on educational web services for implementation personal educational paths of the preferences studying at a basis which mastered the offered educational resources is conducted.

The relevance of a research is caused need of information technology development of expertize and sorting of content in which participate both experts in the researched data domain, and users, at the same time the alternative is selected by means of collective filtering.

Basis of the considered method of collaborative filtering is the assumption that similar educational resources will be useful to similar users of web service. Input information for this method is the matrix the student – the educational resource in which the user estimates are reflected, all estimates of one user represent a vector. The cosine measure allows to estimate similarity of vectors on value of a cosine of the angle in between.

The model is applied to enhancement of technologies of independent operation of students for implementation of personal potential of the student in case of implementation of web portals of educational resources and the organization distant educations.

This research is funded by RFBR, grant 15-07-04102.

- [1] *Lipko Yu.* Research of methods of simulation of recommendatory web services of a choice of educational resources for implementation of the personal educational paths studying at a basis of collaborative filtering // *Izvestiya UFU. Tekhnicheskiye nauki, Rostov-on-Don: Publisher UFU, 2016. — p. 115–122.*

Мера TF-IDF, сила связи слов и формирование единиц представления знаний в открытых тестах

Михайлов Дмитрий Владимирович^{1*} mdv74@list.ru

*Козлов Александр Павлович*¹ caleo@yandex.ru

*Емельянов Геннадий Мартинович*¹ Gennady.Emelyanov@novsu.ru

¹Великий Новгород, Новгородский государственный университет

Разработка открытых тестов предполагает накопление и систематизацию экспертных знаний, исходно представляемых текстами естественного языка (ЕЯ). Источником знаний при этом будут публикации ведущих школ по соответствующей проблематике. В данной работе совместным использованием разбиения слов на классы по значению меры TF-IDF и оценки силы их связи решается задача выделения составляющих образа исходной фразы (СОИФ) в текстах, отбираемых в тематический корпус. Одним из базовых здесь служит предположение о наибольшей значимости связей слов, близких максимальному (наиболее уникальные термины) и среднему (общая лексика, обеспечивающая перифразирование, и термины-синонимы) значениям TF-IDF. При этом сами связи расширяются от традиционных биграмм до трёх и более элементов для выделения СОИФ в виде сочетаний связанных по смыслу слов (с привлечением базы известных синтаксических отношений и без неё). С целью более полного описания выделяемого в текстах фрагмента экспертного знания вводятся в рассмотрение совокупности исходных фраз, взаимно эквивалентных либо дополняющих друг друга по смыслу и представляющих единый образ. По сравнению с поиском СОИФ на готовом синтаксически размеченном текстовом корпусе, охватывающем весь заданный ЕЯ, предложенный метод позволяет в среднем в 17 раз сократить выход фраз, не релевантных исходной ни по описываемому фрагменту знания, ни по ЕЯ-формам его выражения [1].

Работа поддержана РФФИ (проект № 16-01-00004) 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 анализом связей слов в составе n -грамм // Компьютерная оптика, 2017. Т. 41, №3. С.461-471. <http://www.computeroptics.smr.ru/KO/PDF/K041-3/410320.pdf>

TF-IDF, coupling strength of words and formation of units for knowledge representation in open tests

Mikhaylov Dmitry^{1*}

mdv74@list.ru

*Kozlov Aleksandr*¹

caleo@yandex.ru

*Emelyanov Gennady*¹

Gennady.Emelyanov@novsu.ru

¹Veliky Novgorod, Yaroslav-the-Wise Novgorod State University

Development of open form tests requires the accumulating and ordering an expert knowledge initially represented by natural-language (NL) texts. The most natural knowledge source here will be the papers of highestrank scholars in appropriated topical area. In current paper the problem of revelation of initial phrase's image constituents (IPhICs) is solved by estimation of coupling strength of words together with classifying them by TF-IDF metrics relatively to texts selected to topical corpus. One of the base assumptions here is about the most significance of links for words closest to maximal and average values of TF-IDF. These words are correspond to the most unique terms and, respectively, to terms which have synonyms together with general vocabulary by means of which paraphrasing can be carry out. Herewith the links themselves are extended from traditional bigrams to three and more elements for revelation of IPhICs in a form of combinations of related words. The variants of link revelation with and without application of base of known syntactical relations are considered here. To describe more completely the fragment of revealed knowledge the sets of initial phrases mutually equivalent or complementary by sense and related to the same image are entered into consideration. In comparison with the search of IPhICs on a syntactically marked corpus covering all given NL the offered method can reduce, on average, by 17 times the output of phrases which are not relevant to initial ones neither at described knowledge fragment nor at its expression NL-forms [1]. This research is funded by the Russian Foundation for Basic Research, grant 16-01-00004, and the Ministry of Education and Science of Russia (the basic part of the state task).

- [1] *Mikhaylov D., Kozlov A., and Emelyanov G.* 2017. An approach based on analysis of n -grams on links of words to extract the knowledge and relevant linguistic means on subject-oriented text sets. *Comp. Opt.* 41(3): 461-471. <http://www.computeroptics.smr.ru/KO/PDF/KO41-3/410320.pdf>

Использование структуры научных документов в задачах обработки и анализа текстовой информации

Мохов Андрей Сергеевич^{1*}

asmokhov@mail.ru

*Толчеев Владимир Олегович*¹

tolcheevvo@mail.ru

¹Москва, Московский энергетический институт (НИИУ)

В настоящее время при проведении информационного поиска и автоматического аннотирования успешно используется информация о структуре документа, в частности, о значимости того или иного фрагмента. Для библиографических описаний (БО) научных статей, рассматриваемых в данной работе, такими фрагментами являются названия, аннотации и ключевые слова. Нами решается задача разработки классификаторов, способных улучшить точность группирования таких текстов за счет использования информации о наличии терминов в различных разделах БО.

Для описания текстового документа применяется частично структурированная модель, состоящая из трех столбцов, каждый из которых показывает вес термина в названии, аннотации и ключевых словах.

Для классификации используются профильные методы, в которых строятся профили классов, состоящие из наиболее важных классовообразующих терминов [1]. В работе рассматриваются три базовых профильных метода, которые построены на основе хиквадрат критерия, критерия взаимной информации и эвристического коэффициента Жаккара, а также предложенные авторами комбинации этих критериев (алгоритмы UNI5 и UNI6) [1]. Учет структуры документа заключается во взвешивании терминов в трех разделах БО с помощью различных подходов и суммировании получаемых значений.

Для проведения экспериментов сформированы по 20 экзаменационных и обучающих выборок, которые состоят из библиографических описаний научных статей. Приводятся результаты сравнения точностных характеристик известных профильных методов и классификаторов, предложенных в работе.

[1] *Мохов А. С., Толчеев В. О.* Способы учета структуры научных документов в задачах обработки и анализа текстовой информации // Информационные технологии, Москва: №5, 2016. — С. 332–339.

Usage of the structure of scientific documents in tasks of processing and analysis of text information

*Mokhov Andrey*¹*

asmokhov@mail.ru

*Tolcheev Vladimir*¹

tolcheevvo@mail.ru

¹Moscow, Moscow Power Engineering Institute (NRU)

Currently, an information about the structure of the document, in particular, about the importance of a different document parts, is successfully used in tasks of information retrieval and automatic annotation. In this work we discuss bibliographic descriptions (BD) of scientific articles. For BD such fragments are titles, abstracts and keywords. We solve the problem of developing classifiers that can improve the accuracy of grouping such texts by using the information on the availability of terms in different sections of the BD.

For describing a text document we use partly structured model, consisting of three columns, each of which shows the weight of the term in the title, abstract and keywords.

For classification we use profile methods that build profiles of classes, consisting of the most important terms [1]. The paper discusses three basic profile methods, based on the Chi-square criterion, the criterion of mutual information and heuristic Jaccard coefficient and the proposed combination of these criteria (algorithms UNI5 and UNI6) [1]. Based on the structure of the document we weight terms in three parts of BD through a different approaches and summing the obtained values.

To perform the research we generated 20 examination and 20 training samples, which consist of bibliographic descriptions of scientific articles. The results of comparing the accuracy characteristics of known profile methods and well-known classifiers are considered in the article.

- [1] *Mokhov A., Tolcheev V.* The development of high-precision classification methods for bilingual text documents // Information technologies, Moscow: #5, 2016. — pp. 332–339.

Метод семантических ядер для классификации публичных чатов мессенджера Telegram

Новиков Даниил Дмитриевич^{1,2}

diggy-work@yandex.ru

¹Москва, Московский энергетический институт (НИУ)

²Москва, ООО «Комбот»

В последние годы наблюдается быстрый рост пользовательской базы мессенджеров, который в 2015 году превысил рост социальных сетей. Благодаря данному факту появилось множество специфических задач, связанных с проведением обработки и анализа коротких текстовых сообщений. Такие вопросы оказались актуальными и для мессенджера Telegram.

При использовании Telegram у пользователя часто возникает необходимость поиска чата на желаемую тему, но на данный момент в экосистеме мессенджера не существует инструментов, решающих данную проблему автоматически. Для решения данной проблемы в работе решается задача классификации чатов. Предлагается метод семантических ядер — новый способ решения данной задачи. В нём для проведения классификации массивы текстовых сообщений из каждого чата сначала преобразуются в упорядоченные множества слов, ранжированных по частоте употребления. После этого происходит получение семантических ядер путём вычитания множества терминов общей лексики из множества терминов исследуемого чата. Далее из семантического ядра исследуемого чата поочерёдно и независимо вычитаются ядра опорных чатов (эталонные ядра, соответствующие тематическим классам). В конце для каждой такой пары рассчитывается специальная мера семантической близости двух чатов, на основе которой чату выдаётся метка о принадлежности конкретной тематике.

В ходе эксперимента классифицируются 165 чатов по 8 тематикам. Для оценки качества созданного классификатора рассчитывается полнота, точность и доля правильных ответов. Результаты работы алгоритма анализируются в целом. Приводятся характеристики, важные для практического применения.

- [1] *Новиков Д. Д.* Метод семантических ядер для классификации публичных чатов мессенджера Telegram // Машинное обучение и анализ данных, 2017.

Semantic kernels method for Telegram messenger public chats classification

Novikov Daniil^{1,2}

`diggy-work@yandex.ru`

¹Moscow, National Research Nuclear University MEPhI

²Moscow, Combot LLC

In recent years the extreme growth of messengers has been noticed — in 2015 it overcame the growth of social networks. As a result, many specific text analysis tasks have appeared. This issues turned out to be relevant for Telegram messenger too.

To solve this problem, classification task is considered. In this paper author presents a new approach to the text classification — semantic kernels method. According to this algorithm, message arrays are transformed into ordered sets of words ranged by frequency of use. After that, semantic kernels are computed by subtracting a common speech set of terms from a terms set of analyzed chat. Next, kernels of support chats (gold standard kernels that are corresponded to topic classes) are subtracted sequentially and independently from the analyzed chat. After all, for each of these pairs a special semantic similarity measure is calculated. Value of this measure tells which label should be given to a certain chat.

During the experiment, 165 chats are classified by 8 topics. For quality evaluation, precision, recall and accuracy are computed. Performance of the algorithm is analyzed in general. Also, some of its important characteristics are described.

- [1] *Novikov D. D.* Semantic kernels method for Telegram messenger public chats classification // Journal of Machine Learning and Data Analysis, 2017.

Парадигма смыслового проектирования онтологий

Рогозов Юрий Иванович

rogozov@tti.sfedu.ru

Таганрог, Южный федеральный университет

Известно, что процесс создания онтологий является познавательным процессом. Познавательный процесс создания онтологий в виде совокупности конкретных понятий и отношений между ними для каждой из предметных областей, является эмпирическим и практически не формализован [1]. Суть предлагаемой идеи заключается в том, что в основу онтологии положены не конкретные понятия и отношения между ними, а средства создания конкретных понятий. Каждое понятие представляется в виде познавательного средства, в котором происходит процесс самоорганизации конкретных понятий и их взаимопроникновение. В основе предлагаемой парадигмы лежат несколько предположений. Первое предположение гласит, что организованная определенным образом совокупность системных свойств может самоорганизоваться в системное средство, в котором могут самопорождаться конкретные понятия [1]. К таким системным свойствам предлагается отнести: рефлексию; методологию; логику; эмерджентность. Второе устанавливает, что рефлексия, методология и логика определяют пространственное представление понятия, а эмерджентность временную составляющую процесса воспроизводства. Под рефлексией предлагается понимать строение способа познания (логика рефлексии), которая является той субстанцией, в смысле Аристотеля («губкой без атрибутов»), которую необходимо заполнить определенными атрибутами, чтобы она стала конкретным средством. Методология выполняет функцию организации единства. Логика организует органическое единство конкретных средств познания. Эмерджентность организует системное средство воспроизводства. Предлагаемая парадигма решает междисциплинарную проблему, т.к. при создании объекта устанавливается связь не между понятиями различных предметных областей, а между правилами или строениями правил организации правил.

Работа поддержана грантом РФФИ № 15-07-04429.

- [1] *Рогозов Ю. И.* Парадигма смыслового проектирования системных объектов // Труды ИСА РАН, Т.67, №3, 2017. — с. 41-53.

The paradigm of the semantic design of ontologies

Rogozov Yury

rogozov@tti.sfedu.ru

Taganrog, Southern Federal University

It is known that the process of creating the ontologies is a cognitive process. The cognitive process of creating the ontologies in the form of a set of concrete concepts and relations between them for each of the subject area is empirical and practically not formalized process [1]. The essence of the proposed idea is that ontology is based not on specific concepts and relations between them, but on the means of creating concrete concepts. Each concept is represented as a cognitive means, in which the process of self-organization of specific concepts and their interpenetration is taking place. There are several assumptions at the core of the proposed paradigm. The first assumption is that a set of system properties organized in a certain way can self-organize into a system tool in which specific concepts can be self-generated [1]. It is proposed that such system properties include: reflection; methodology; logics; emergence. The second assumption establishes that reflection, methodology and logic determine the spatial representation of the concept, and emergence determines the temporal component of the reproduction process. Reflexion is proposed to understand as the process of creating the structure of the cognition method or reality perception (the logic of reflection), which is substance, in the Aristotle sense (“sponge without attributes”), which must be filled with certain attributes to make it a concrete tool. The methodology performs the function of organizing the unity between the individual cognitive means. Logic organizes the organic unity of concrete means of cognition. Emergence organizes the organic unity of all these means in a systemic means of reproducing the specific physical forms of target objects. The proposed paradigm solves the interdisciplinary problem, because when creating an object, a relationship is established not between the concepts of different subject areas, but between the rules or structures of rules for organizing the rules.

The study was supported by RFBR, project No. 15-07-04429.

- [1] *Rogozov Yu. I.* The paradigm of semantic system objects design // Proceedings of the ISA RAS V.64 No.3, 2017. — pp.41-53.

Классификация физической активности человека с помощью локальных аппроксимирующих моделей

Аникеев Дмитрий Александрович^{1*}

dmitriy.anikeev@phystech.edu

*Пенкин Григорий Олегович*²

penkin.gr@gmail.com

Стрижов Вадим Викторович^{1,3}

strijov@phystech.edu

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Московский государственный университет

им. М. В. Ломоносова

³Москва, Федеральный исследовательский центр «Информатика и управление» РАН

Рассматривается задача классификации временных рядов акселерометра мобильного телефона. Классу физической активности соответствует сегмент временного ряда. Такому сегменту сопоставляется его локальное описание. В качестве локального описания используются параметры сплайна (алгебраического, сглаживающего, адаптивного регрессионного или сплайна с динамическими узлами), аппроксимирующего временной ряд. Логистическая регрессия отображает вектор признаков, состоящий из коэффициентов сплайна, в метку класса.

Решается проблема использования значения параметров локальных аппроксимирующих моделей. Точность классификации элементов выбранного признакового пространства зависит от способа построения этого пространства. Вычислительный эксперимент находит оптимальные параметры аппроксимации и параметры модели классификации согласно максимуму правдоподобия логистической модели классификации.

Работа поддержана РФФИ, проект 16-07-01158.

- [1] *Аникеев Д. А., Пенкин Г. О., Стрижов В. В.* Классификация физической активности человека с помощью аппроксимирующих моделей // Информатика и ее применение, 2018 (подготовлено к печати). <http://www.ipiran.ru/journal/issues/>.

Local approximation for human physical activity classification

Anikeyev Dmitriy^{1*}

dmitriy.anikeyev@phystech.edu

*Penkin Grigory*²

penkin.gr@gmail.com

Strijov Vadim^{1,3}

strijov@phystech.edu

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

²Moscow, Lomonosov Moscow State University

³Moscow, Federal Research Center “Computer Science and Control”
of RAS

The research is devoted to the time series classification. The time series is measured by the accelerometer of a wearable device. A class of physical activity is defined by the feature description of a time segment. To construct this description the authors propose to use parameters of various approximation splines (algebraic, smoothing, adaptive regression or spline with dynamic nodes). The logistic regression is used as a classifier. It delivers desired quality of the activity recognition

The authors analyse the space of the local approximation parameters. Classification accuracy depends on the method of this space construction. The computational experiment finds the optimal approximation parameters and parameters of the classifier. This research is funded by RFBR, grant 16-07-01158.

- [1] *Anikeyev D., Penkin G., Strijov V.* Classification of human physical activity using approximate models. // Informatics and Application, 2018 (prepared for publication). <http://www.ipiran.ru/english/journal/index.asp>.

Локализация источников мозговой активности по МЭГ данным в исследованиях семантического прайминга

Буторина Анна Валерьевна^{1*}

armature@yandex.ru

*Павлова Анна Александровна*¹

anne.al.pavlova@gmail.com

¹Москва, МЭГ-центр, МГППУ

Цель данного исследования заключалась в том, чтобы показать, что речевой стимул-праймер вызывает быструю активацию распределенной семантической сети мозга. В магнитоэнцефалографическом (МЭГ) исследовании мы просили испытуемых назвать глагол, ассоциирующийся с предъявленным существительным словом. Предъявляемые слова-праймеры отличались семантической нагрузкой: в одном условии ассоциацию назвать было легко, в другом - затруднительно. МЭГ - это неинвазивный метод картирования активности мозга, который обладает хорошим пространственным разрешением. Анализ МЭГ данных состоял из двух этапов. Во-первых, анализ источников, призванный характеризовать пространственно-временную динамику, лежащую в основе обработки существительных в задаче генерации подходящих глаголов. Кортикальные источники вызванных ответов были смоделированы с использованием метода linear L2-minimum norm estimation. Вторая часть анализа была направлена на выявление предполагаемого влияния силы связи «существительное-глагол» на мозговую активность. Этот шаг был выполнен с помощью байесовского алгоритма локализации источников Multiple Sparse Priors. Теория случайных полей была использована для оценки статистической значимости результатов. Значимые различия в обработке слова-праймера наблюдались уже через 250-400 мс после начала стимула. Так как данное время совпадает с компонентом N400 «ответственным» за семантическую обработку слова, можно предположить, что поиск целевого глагола необязательно требует завершения обработки существительного, но частично перекрывается и взаимодействует с его семантическим анализом.

[1] *Butorina A., Pavlova A., Nikolaeva A., Prokofyev A., Bondarev B., Stroganova T.* Simultaneous Processing of Noun Cue and to-be-Produced Verb in Verb Generation Task: Electromagnetic Evidence // *Frontiers in human neuroscience* 11, 2017. — С. 1–14.

MEG source localization in studies of semantic priming

Butorina Anna^{1*}

armature@yandex.ru

*Pavlova Anna*¹

anne.al.pavlova@gmail.com

¹Moscow, MEG-center, MSUPE

The current study was aimed to provide direct evidence that presented cue evoke rapid spread of activation within brain semantic networks. By measuring magnetoencephalography (MEG) from participants involved in verb generation task, we examined whether neural responses evoked by noun cues were modulated by the association strength between a cue and to-be-produced verb. MEG is a non-invasive technique for mapping brain activity and has an excellent spatial resolution; sources can be localized with millimeter precision. The analysis of MEG data had two steps. First, a source analysis intended to characterize the spatial-temporal cortical dynamics underlying noun cue processing in verb generation task. The cortical sources of the evoked responses were modeled by a “depth-weighted” linear L2-minimum norm estimation method. The second part of the analysis was aimed to identify the putative effect of noun-verb association strength on neural activity evoked by noun cue presentation, both in terms of its timing and involved brain regions. This step was done by the means of Multiple Sparse Priors, the Bayesian source inversion algorithm. We used Random Field Theory for statistical analysis. Our main finding is that the strength of noun-verb association affects noun’s processing already at 250–400 ms after stimulus onset, i.e., at the stage of semantic M400 component of noun-evoked response or even earlier. Thus, the retrieval of target verb does not necessarily require the noun’s processing to be completed but partially overlaps and interacts with its semantic analysis.

- [1] *Butorina A., Pavlova A., Nikolaeva A., Prokofyev A., Bondarev B., Stroganova T.* Simultaneous Processing of Noun Cue and to-be-Produced Verb in Verb Generation Task: Electromagnetic Evidence // *Frontiers in human neuroscience* 11, 2017. — p. 1–14.

Автоматическое детектирование событий в долговременных записях ЭЭГ

Кершнер Иван Андреевич^{1*} ivan.kershner@gmail.com
*Комольцев Илья Геральдович*² outaudiofillin@gmail.com
*Обухов Константин Юрьевич*³ 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, Московский физико-технический институт (ГУ)

Предложен новый способ автоматического детектирования событий в сигналах длительного (сутки) мониторинга ЭЭГ, основанный на анализе 3-D хребтов вейвлет спектрограмм этих сигналов. Рассчитывалась гистограмма количества событий N в зависимости от спектральной плотности мощности (PSD) в точках хребта, имеющая максимум N_{max} при некотором значении PSD и спадающая до $N = 1$ при больших значениях PSD. Для выделенных событий N_{max} рассчитывались их частотные, временные и PSD характеристики.

Построена модель логистической регрессии для классификации сонных веретён и эпилептиформной активности среди событий ЭЭГ крыс после черепно-мозговой травмы. Обучение модели проводилось на событиях, классифицированных тремя независимыми экспертами-нейрофизиологами как эпилептиформные и сонные веретёна.

Исследование выполнено за счёт гранта Российского научного фонда (проект № 16-11-10258).

- [1] *Obukhov K., Kershner I., Komoltsev I., Maluta I., Obukhov Yu., Manolova A., Gulyaeva N.* An Approach for EEG of Post Traumatic Sleep Spindles and Epilepsy Seizures Detection and Classification in Rats // Pattern Recognition and Image Analysis, 2017. — Vol. 27, No. 1. — pp. 114-121. <https://doi.org/10.1134/S1054661817010102>.

Approach for EEG of Post Traumatic Sleep Spindles and Epilepsy Seizures Detection and Classification in Rats

Kershner Ivan^{1*}

ivan.kershner@gmail.com

*Komoltsev Ilya*²

outaudiofillin@gmail.com

*Obukhov Konstantin*³

ko.arsenalfc@gmail.com

*Obukhov Yury*²

yuvobukhov@mail.ru

¹Moscow, Kotel'nikov Institute of Radio Engineering and Electronics of RAS

²Moscow, Institute of Higher Nervous Activity and Neurophysiology of RAS

³Moscow, Moscow Institute of Physics and Technology (State University)

The electroencephalographic (EEG) features of post traumatic epilepsy (PTE) are analyzed in the paper. The proposed method allows detection and classification of sleep spindles and epilepsy seizures. The experiments were conducted on a laboratory rats before and after traumatic brain injury (TBI). In the introduction, the details of the experiment along with the information about manual markup are provided. In the first part, the new method of sleep spindles and epilepsy seizures detection is described. The method is based on the analysis of the wavelet spectrogram extrema. Moreover, the described procedure of background extraction and ridge segmentation helps to classify signals as epilepsy seizures and sleep spindles. In the second part, the information about the clustering is given. k-Means clustering of seizures and spindles was performed based on signals power and frequency. The results of the clustering, along with the research of TBI effect on the EEG, are provided in the third part. It was shown that PTE may be considered as the cause of the frequency variance among clusters of sleep spindles and epilepsy seizures.

This research is funded by Russian Science Foundation (project No. 16-11-10258).

- [1] *Obukhov K., Kershner I., Komoltsev I., Maluta I., Obukhov Yu., Manolova A., Gulyaeva N.* An Approach for EEG of Post Traumatic Sleep Spindles and Epilepsy Seizures Detection and Classification in Rats // *Pattern Recognition and Image Analysis*, 2017. — Vol. 27, No. 1. — pp. 114-121. <https://doi.org/10.1134/S1054661817010102>.

Анализ биомедицинских сигналов с нелинейными свойствами на основе энтропийных характеристик

*Манило Людмила Алексеевна*¹

lmanilo@yandex.ru

Немирко Анатолий Павлович^{1*}

apn-bs@yandex.ru

¹ Санкт-Петербург, Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»

Рассматривается метод анализа сигналов с хаотическими свойствами по параметрам аппроксимированной энтропии. Эта характеристика является приближенной оценкой энтропии Колмогорова, и она может быть использована для обнаружения динамических нерегулярностей, связанных как с нелинейными, так и со стохастическими свойствами сигналов. Отличие ее от энтропии Шеннона заключено в механизме формирования ячеек фазового пространства. Выбор каждой точки, в которой находится объект, в качестве центра гиперсфер приводит к более трудоемким вычислениям. Существует быстрый алгоритм вычисления аппроксимированной энтропии. Это позволяет использовать ее для анализа биомедицинских сигналов в реальном времени. Модельные эксперименты показали возможность получения надежных оценок хаотичности даже на коротких выборках данных. Нечувствительность аппроксимированной энтропии к воздействию шума обеспечивается выбором определенного размера ячеек фазового пространства. Предложен набор параметров для оценки степени регулярности в дискретной последовательности отсчетов сигнала. Предложена процедура коррекции, снижающая вероятность ошибки в анализе сигналов с нелинейными свойствами. Эксперименты по распознаванию биомедицинских сигналов (ЭКГ, ЭЭГ) для ряда практических задач показали, что методы анализа аппроксимированной энтропии могут найти применение в системах медицинской диагностики и наблюдения.

Работа поддержана грантами РФФИ № 16-01-00159, 15-07-01790.

- [1] *Манило Л. А., Миганько А. В.* Исследование параметров и свойств аппроксимированной энтропии ЭЭГ в разных стадиях анестезии // Биотехносфера, 2017, 2 (50). — С. 21–26.

Analysis of biomedical signals with nonlinear properties based on entropy characteristics

*Manilo Ludmila*¹

lmanilo@yandex.ru

*Nemirko Anatoly*¹✉

apn-bs@yandex.ru

¹Saint Petersburg, Saint Petersburg Electrotechnical University “LETI”

A method for signal with chaotic properties analysis using the approximated entropy parameters is considered. This feature is an approximate estimate of the Kolmogorov entropy and it can be used to detect dynamic irregularities associated with the nonlinear and stochastic properties of signals. The difference it from Shannon's entropy lies in the mechanism of phase space cells formation. The choice of each point, which is the object, as the hypersphere center leads to more time consuming calculations. However, there is a fast algorithm for computing the approximated entropy. This allows you to use it for the analysis of biomedical signals in real time. Model tests have shown the possibility of obtaining reliable estimates of randomness, even on short data samples. The insensitivity of the approximated entropy to the effects of noise is provided by the selection of a particular cell size of the phase space. A set of parameters for the degree of regularity assessment in the discrete sequence of signal samples is proposed. The correction procedure, which reduces the probability of error in the analysis of signals with nonlinear properties, is proposed. Experiments on the recognition of biomedical signals (ECG, EEG) for a number of practical problems have shown that methods of the approximated entropy analysis can be used in systems for medical diagnosis and monitoring.

This research is funded by RFBR, grants 16-01-00159, 15-07-01790.

- [1] *Manilo L. A., Miganko A. V.* Study of parameters and properties of approximate entropy of EEG at various anesthesia stages // *Biotechnosphere*. - 2017, No 25 (50). — p. 21–26.

Разработка методов и алгоритмов анализа видеоизображений лабораторных животных средствами объектно-ориентированного логического программирования

Морозов Алексей Александрович^{1*}

morozov@cplire.ru

*Сушкова Ольга Сергеевна*¹

o.sushkova@mail.ru

*Полупанов Александр Фёдорович*¹

sashap55@mail.ru

¹Москва, Институт радиотехники и электроники
им. В. А. Котельникова РАН

Рассмотрена проблема анализа видеоизображений лабораторных животных (крыс) средствами объектно-ориентированного логического программирования. Видеоизображения получены в экспериментах по изучению судорожной активности головного мозга. Осуществлялась одновременная запись ЭЭГ животного и видеозапись его поведения. Сопоставление данных ЭЭГ и поведения необходимо, в частности, потому, что резкие движения животного могут приводить к появлению в ЭЭГ-записях сигналов, похожих на эпилептические разряды. Поэтому одной из задач видеонаблюдения является распознавание резких движений и использование этой информации при интерпретации результатов эксперимента. Другой задачей является анализ поведения животного при когнитивных тестах (социальное опознавание и опознавание нового объекта). Существенной особенностью видеозаписей является то, что видеонаблюдение проводилось в том же боксе, где животное живёт на подстилке из опилок. Цвет шерсти животного мало отличается от цвета опилок, поэтому выделение контура животного является сложной задачей. В докладе рассказывается о разработке алгоритмов низкоуровневого анализа видеоизображений, а также логических методов анализа поведения животного. Методы и алгоритмы реализованы в объектно-ориентированном логическом языке Акторный Пролог.

Исследование выполнено за счёт гранта Российского научного фонда (проект № 16-11-10258).

- [1] Морозов А. А., Сушкова О. С. Анализ видеоизображений в реальном времени средствами языка Акторный Пролог // Компьютерная оптика. — 2016. — Т. 40, № 6. — С. 947–957. <http://www.computeroptics.smr.ru/KO/PDF/KO40-6/400623.pdf>.

Development of methods and algorithms based on object-oriented logic programming for video monitoring of laboratory animals

*Morozov Alexei*¹*

morozov@cplire.ru

*Sushkova Olga*¹

o.sushkova@mail.ru

*Polupanov Alexander*¹

sashap55@mail.ru

¹Moscow, Kotel'nikov Institute of Radio Engineering and Electronics of RAS

The problem of analysis of video images of laboratory animals (rats) by the means of object-oriented logic programming is considered. The videos are produced in neurophysiologic experiments on the study of convulsive electrical activity of the brain cortex. Videos of behaviour of laboratory rats were recorded simultaneously with electroencephalograms (EEG) in the animals. A comparison of EEG data with the behaviour of animals is necessary because sharp motions of the animals can result in EEG signals that are very similar to the epileptic discharges. Thus, the first task of the video monitoring is recognition of the sharp motions of the animals and using this information for proper interpretation of the results of the experiments. The second task of the video monitoring is analysis of behaviour of animals in cognitive testing (social recognition and recognition of a new object). Essential feature of the video records is in that they are conducted in the same cage where the animal lives, that is, the background of the cage is sawdust. The colour of the animals is about the same as the colour of the sawdust thus the detection of the contour of an animal is not a simple problem. In the report, development of low-level algorithms for video analysis as well as logical methods for the analysis of animal behaviour is discussed. The methods and algorithms are implemented in the Actor Prolog object-oriented logic language.

This research is funded by Russian Science Foundation (project No.16-11-10258).

- [1] *Morozov A., Sushkova O.* Real-time analysis of video by means of the Actor Prolog language // *Computer Optics*. — 2016. — V.40, No.6. — PP.947–957. <http://www.computeroptics.smr.ru/KO/PDF/K040-6/400623.pdf>.

Выбор признаков в задаче декодирования сигналов ECoG

Мотренко Анастасия Петровна
anastasiya.motrenko@phystech.edu

Москва, Московский физико-технический институт (ГУ)

Работа посвящена проблеме разработки математических методов, лежащих основе нейро-компьютерных интерфейсов. Решается задача прогнозирования движений руки по измерениям кортикальной активности. Признаковое описание для каждого отсчета времени лежит в пространственно-частотно-временном диапазоне и включает сигналы, снятые с электродов, и их спектральные характеристики. Представление данных в виде многомодальных матриц широко используется в анализе биомедицинских сигналов.

Так как измерений электрокортикограмм сильно коррелированы для соседних электродов и отсчетов времени, выбор признаков необходим для построения устойчивой модели прогнозирования. Предлагается многомодальная формулировка фильтрационного метода выбора признаков, основанного на решении задачи квадратичного программирования (Quadratic Programming Feature Selection, QPFS). Задача выбора признаков формулируется в виде задачи минимизации функционала, уравнивающего сходство между признаками и релевантность признаков решаемой задаче. Для адаптации метода QPFS к многомодальным данным вводится отдельный функционал качества для каждой модальности признакового описания. Такой подход снижает размерность решаемой оптимизационной проблемы. Приводится сравнение исходного QPFS и его многомодальной модификации при решении задачи восстановления траектории движений руки. Продемонстрировано, что предложенная модификация позволяет увеличить размерность исходного множества признаков при сопоставимом качестве решения задачи регрессии.

Работа поддержана грантом РФФИ № 16-07-01158.

- [1] *Motrenko A, Strijov V.* Extracting fundamental periods to segment biomedical signals // *Journal of Biomedical and Health Informatics*, 2016, Vol. 20, No. 6, 1466–1476.

Multi-way Feature Selection for ECoG-based Brain-Computer Interface

Anastasia Motrenko

anastasiya.motrenko@phystech.edu

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

We address the problem of designing Brain-Computer Interfaces. The task is to predict hand trajectories from the voltage time series of cortical activity. Feature description of a each point resides in spatial-temporal-frequency domain, which includes the voltage time series and their spectral characteristics. Such multi-way representation is actively used in biomedical applications. Unfolding multi-way data into flat matrices might lead to neglecting important dependencies present in the unfolded dimension of the multi-way data.

Contrarily, multi-way approaches preserve the data structure and improve regression quality.

Feature selection is crucial for adequate solution of this regression problem, since electrocorticographic data is highly dimensional and the measurements are correlated both in time and space domains. We propose a multi-way formulation of a recent approach to filtering feature selection, Quadratic Programming Feature Selection (QPFS). Filtering methods separate feature selection from regression model training. This makes them computationally efficient even for large dimensionalities. Since the original QPFS ignores multi-way structure of the data, we propose a multi-way extension of QPFS. We introduce a separate feature similarity matrix for each modality of the feature description. This reduces dimensionality of optimization problem, which makes the proposed approach applicable even in case of high dimensionality. We compare the original and multi-way QPFS applied to trajectory reconstruction problem and show that proposed modification applies to higher dimensionalities without loss in quality.

This work was supported by the Russian Foundation for Basic Research, grant 16-07-01158.

- [1] *Motrenko A, Strijov V.* Extracting fundamental periods to segment biomedical signals // *Journal of Biomedical and Health Informatics*, 2016, Vol. 20, No. 6, 1466–1476.

Mathbrain.ru — облачный сервис для обработки данных магнитной энцефалографии

Оплачко Екатерина Сергеевна^{1*} oplachkoe@gmail.com

*Рыкунов Станислав Дмитриевич*¹ rykunov@psn.ru

*Устинин Михаил Николаевич*¹ ustinin@impb.ru

¹Пушино, Институт математических проблем биологии РАН

MathBrain является облачным приложением, распространяемым по модели «Программное обеспечение как сервис» (Software as a Service). Это приложение предоставляет доступ к нескольким алгоритмам анализа данных многоканальной энцефалографии. Спектральные методы включают в себя прямое и обратное преобразование Фурье, а также количественный анализ. Статистические методы включают метод главных компонент и метод независимых компонент. Карты поля элементарных составляющих могут быть использованы для решения обратной задачи магнитной энцефалографии, результат которого отображается на магнитно-резонансной томограмме. Приложение разработано для изучения человеческого мозга, в том числе, специалистами, не имеющими специальной математической подготовки.

Работа поддержана грантами РФФИ № 16-07-00937, № 16-07-01000, № 17-07-00677, № 17-07-00686, № 17-29-02178, программой фундаментальных исследований Президиума РАН № I.33П.

- [1] *Рыкунов С. Д., Устинин М. Н., Полянин А. Г., Сычев В. В., Линас Р. Р.* Комплекс программ для расчета парциальных спектров головного мозга человека // Математическая биология и биоинформатика, Пушино, 2016. — Т.11 № 1. С. 127-140 doi: 10.17537/2016.11.127 http://www.matbio.org/2016/Rykunov_11_127.pdf.

Mathbrain.ru — a cloud service for magnetic encephalography data processing

Oplachko Ekaterina^{1*}

oplachkoe@gmail.com

*Rykunov Stanislav*¹

rykunov@psn.ru

*Ustinin Mikhail*¹

ustinin@impb.ru

¹Pushchino, Institute of Mathematical Problems of Biology RAS

MathBrain is a cloud-based application, distributed under Software as a Service model. This application provides access to several algorithms of the multichannel encephalography data analysis. Spectral methods include direct and inverse Fourier transforms and quantitative analysis. Statistical methods involve principal component analysis and independent component analysis. The field maps of elementary components can be used to solve the magnetic encephalography inverse problem and to display the result at the magnetic resonance image. The application is designed to be used in human brain studies without mathematical training.

The research was partly supported by the RFBR (grants 16-07-00937, 16-07-01000, 17-07-00677, 17-07-00686, 17-29-02178) and by the Program of the Presidium of RAS I.33P.

- [1] *Rykunov S. D., Oplachko E. S., Ustinin M. N., Llinás R. R.* Methods for magnetic encephalography data analysis in MathBrain cloud service // *Mathematical biology and bioinformatics*, Pushchino, 2017. 12(1):176-185 doi: 10.17537/2017.12.176 http://www.matbio.org/2017/Rykunov_12_176.pdf.

Нахождение парциальных спектров разделов головного мозга методом анализа в пространстве частота-паттерн

Рыкунов Станислав Дмитриевич^{1*} rykunov@psn.ru
*Сычев Вячеслав Викторович*¹ sychyov@yahoo.com
*Молчанова Дина Альбертовна*¹ dm@pochtamt.ru
*Устинин Михаил Николаевич*¹ ustinin@impb.ru

¹Пушино, Институт математических проблем биологии РАН

Разработан новый метод изучения активности мозга человека, направленный на определение спектральных характеристик, специфичных для различных областей и структур человеческого мозга. Этот метод объединяет два различных вида информации: 1) анатомическую информацию, полученную из магнитно-резонансной томограммы (МРТ); 2) информацию о мозговой активности, полученную многоканальным магнитным энцефалографом. Для определения пространственного положения источников активности применяется метод функциональной томографии. Производится сегментация МРТ и выделяются области интереса. Наложение этих областей на функциональную томограмму дает набор источников поля, соответствующих выбранной области. Частоты и коэффициенты Фурье этих источников формируют парциальный спектр исследуемой области или структурного элемента человеческого мозга. Для проверки метод был применен к записям спонтанной активности двух групп испытуемых: 1) здоровые испытуемые; 2) испытуемые с определенными патологиями (рассеянный склероз и невралгии различного генеза). Полученные результаты показывают хорошую корреляцию с общепринятыми представлениями о локализации источников такой активности.

Работа поддержана грантами РФФИ № 16-07-00937, № 16-07-01000, № 17-07-00677, № 17-07-00686, № 17-29-02178, программой фундаментальных исследований Президиума РАН № I.33П.

[1] *Рыкунов С. Д., Устинин М. Н., Полянин А. Г., Сычев В. В., Линас Р. Р.* Комплекс программ для расчета парциальных спектров головного мозга человека // Математическая биология и биоинформатика, Пушино, 2016. — Т.11 №1. С. 127-140 doi: 10.17537/2016.11.127

Estimation of the brain compartments partial spectra by frequency-pattern analysis

Rykunov Stanislav^{1*}

rykunov@psn.ru

*Sychev Vyacheslav*¹

sychyov@yahoo.com

*Molchanova Dina*¹

dm@pochtamt.ru

*Ustinin Mikhail*¹

ustinin@impb.ru

¹Pushchino, Institute of Mathematical Problems of Biology RAS

A new method to study human brain activity was developed. This method aims to determine spectral characteristics, specific to various regions and structures of the human brain. Said method combines two completely different sources of information: 1) anatomical information, obtained from magnetic resonance imaging (MRI); 2) information about human brain activity, obtained by multichannel magnetic encephalograph. To find activity sources from magnetoencephalogram functional tomography method is used. Functional tomography finds corresponding unique spatial location for each coherent elementary oscillation. Spatial information about regions of interest (ROI) is extracted from MRI either by semi-automatic segmentation, or by direct selection. Combining these ROIs with functional tomogram one can obtain set of the field sources, corresponding to selected ROI. Frequencies and Fourier coefficients of these sources form partial spectrum of the studied region or structural element of the human brain. For verification, the method was applied to the records of the human brain spontaneous activity obtained from healthy subjects and from subjects with certain pathologies (multiple sclerosis and neuralgias of different genesis). The results obtained show good correlation with the generally accepted ideas about the localization of sources of such activity.

The research was partly supported by the RFBR (grants 16-07-00937, 16-07-01000, 17-07-00677, 17-07-00686, 17-29-02178) and by the Program of the Presidium of RAS I.33P.

- [1] *Rykunov S. D., Ustinin M. N., Polyinin A. G., Sychev V. V., Llinás R. R.* Software for the partial spectroscopy of the human brain // *Mathematical biology and bioinformatics*, Pushchino, 2016. 11(1):127-140 doi: 10.17537/2016.11.127

Методы машинного обучения в решении задачи мониторинга выведения тераностических флуоресцентных нанокомпозитов из организма

<i>Сарманова Ольга Эдуардовна</i> ¹	helga-sharman@rambler.ru
<i>Буриков Сергей Алексеевич</i> ^{1,2}	sergey.burikov@gmail.com
<i>Доленко Сергей Анатольевич</i> ²	dolenko@srd.sinp.msu.ru
<i>Исаев Игорь Викторович</i> ^{2*}	isaev_igor@mail.ru
<i>Светлов Всеволод Александрович</i> ²	svetlov.vsevolod@gmail.com
<i>Лаптинский Кирилл Андреевич</i> ^{1,2}	laptinskij@physics.msu.ru
<i>Доленко Татьяна Альдефонсовна</i> ^{1,2}	tdolenko@mail.ru

¹Москва, МГУ им. М.В. Ломоносова, Физический факультет

²Москва, МГУ им. М.В. Ломоносова, НИИЯФ им. Д.В. Скобельцына

В настоящее время в биомедицине чрезвычайно актуальна разработка новых наноматериалов, используемых одновременно для диагностики и лечения заболеваний. При создании таких агентов необходимо контролировать выведение их из организма.

В работе представлены результаты применения методов машинного обучения для мониторинга в моче выведенных из организма тераностических нанокомпозитов (углеродных точек, покрытых сополимером и фоллиевой кислотой) и их компонентов по спектрам флуоресценции. Задача решалась как задача кластеризации (с использованием k-means и разрабатываемого авторами алгоритма адаптивного построения иерархических нейросетевых классификаторов - ИНК) и как задача классификации (с помощью нейронных сетей). Ни одно из разбиений на кластеры не показало чувствительности к типам содержащихся в суспензии наночастиц. Наилучшие результаты решения задачи классификации обеспечил перцептрон с 8 нейронами в единственном скрытом слое, натренированный на существенных входных признаках, выделенных кросс-корреляцией. Процент правильного распознавания, усредненный по всем 5 классам, составил 72.3%.

Работа выполнена за счет грантов РФФИ №15-07-08975-а (алгоритм ИНК) и РФФИ №17-12-01481 (все остальные работы).

- [1] *Сарманова О. Э. и др.* Методы машинного обучения в решении задачи мониторинга выведения тераностических флуоресцентных нанокомпозитов из организма // Машинное обучение и анализ данных, 2017.

Machine learning methods for the purpose of monitoring of the excretion of theranostic fluorescent nanocomposites out of the organism

<i>Sarmanova Olga</i> ¹	helga-sharman@rambler.ru
<i>Burikov Sergey</i> ^{1,2}	sergey.burikov@gmail.com
<i>Dolenko Sergey</i> ²	dolenko@srd.sinp.msu.ru
<i>Isaev Igor</i> ^{2*}	isaev_igor@mail.ru
<i>Svetlov Vsevolod</i> ²	svetlov.vsevolod@gmail.com
<i>Laptinskiy Kirill</i> ^{1,2}	laptinskiy@physics.msu.ru
<i>Dolenko Tatiana</i> ^{1,2}	tdolenko@mail.ru

¹Moscow, Lomonosov MSU, Department of Physics

²Moscow, Lomonosov MSU, Skobeltsyn Institute of Nuclear Physics

At present time the development of new nanomaterials, that are used for diagnostics and medical treatment simultaneously, is utterly relevant in biomedicine. During the creation of such agents one has to control their excretion out of the body.

In this study the results of the application of machine learning methods for monitoring of the excreted theranostic nanocomposites (carbon dots, covered by copolymer and folic acid) and their components by their fluorescence spectra in urine are presented. The problem was solved as a clusterization problem (by k-means and algorithm of adaptive construction of hierarchical neural classifiers – HNC – which was developed by the authors) and as a classification problem (by neural networks (NN)). None of the clusterings revealed sensitivity to types of nanoparticles contained in the suspension. The best results of the solution of the classification problem were provided by perceptron with 8 neurons in the hidden layer, trained on the set of significant features selected by cross-correlation. Percentage of correct recognition averaged over all five classes was 72.3%.

This study has been performed at the expense of the following foundations: Russian Foundation for Basic Research (grant no.15-07-08975-a) (HNC algorithm) and Russian Science Foundation (grant no.17-12-01481 (all other studies).

- [1] *Sarmanova O. E. et al.* Machine learning methods for the purpose of monitoring of the excretion of theranostic fluorescent nanocomposites out of the organism // *Journal of Machine learning and data analysis*, 2017.

Исследование специфичности признаков раннего паркинсонизма, полученных методом анализа всплескообразной электрической активности коры мозга

Сушкова Ольга Сергеевна^{1*}

o.sushkova@mail.ru

*Морозов Алексей Александрович*¹

morozov@cplire.ru

*Габова Александра Васильевна*²

agabova@yandex.ru

¹Москва, Институт радиотехники и электроники
им. В. А. Котельникова РАН

²Москва, Институт высшей нервной деятельности и
нейрофизиологии РАН

Разработан новый метод анализа сигналов на основе вейвлет-анализа, ROC-анализа и непараметрической статистики для детального изучения частотно-временной динамики электрической активности коры головного мозга. Идея метода заключается в том, что электроэнцефалограмма (ЭЭГ) рассматривается как набор «всплесков»; при этом мы рассматриваем всплеск не как особую разновидность сигналов ЭЭГ, а как типичный компонент ЭЭГ. Всплески определяются как локальные максимумы на вейвлет-спектрограмме ЭЭГ. Исследуются следующие параметры всплесков: ведущая частота, длительность, ширина полосы частот, количество всплесков в секунду, мощность. Анализируется степень отличия группы пациентов на первой стадии болезни Паркинсона от здоровых добровольцев в пространстве этих параметров. Для этого используется ROC-анализ. Исследуется функциональная зависимость AUC (площади под ROC-кривой) от значений границ диапазонов этих параметров. Используя этот метод, мы выявили три частотных диапазона, где были обнаружены отличия между группой пациентов и здоровыми добровольцами. В докладе рассказывается об исследовании специфичности этих признаков раннего паркинсонизма.

Работа осуществлена при финансовой поддержке программы Президиума РАН № I.33П.

- [1] *Sushkova O. S., Morozov A. A., Gabova A. V.* Data Mining in EEG Wave Trains in Early Stages of Parkinson's Disease // *Advances in Soft Computing. LNCS.* — Cham: Springer, 2017. — P. 403–412. https://link.springer.com/chapter/10.1007/978-3-319-62428-0_32.

An investigation of specificity of features of early stages of Parkinson's disease obtained using the method of cerebral cortex electrical activity analysis based on wave trains

*Sushkova Olga*¹*

`o.sushkova@mail.ru`

*Morozov Alexei*¹

`morozov@cplire.ru`

*Gabova Alexandra*²

`agabova@yandex.ru`

¹Moscow, Kotel'nikov Institute of Radio Engineering and Electronics of RAS

²Moscow, Institute of Higher Nervous Activity and Neurophysiology of RAS

A new method of signal analysis based on wavelet analysis, ROC-analysis, and non-parametric statistics for a detailed investigation of the time-frequency dynamics of the electrical activity of the cerebral cortex is developed. The idea of the method is in that the electroencephalogram (EEG) is considered as a set of wave trains. We consider the wave train as a typical component of the EEG, but not as a special kind of EEG signals. Wave trains are detected as local maxima in the wavelet spectrogram of the EEG. The following parameters of the wave trains are accounted: the frequency, the duration, the bandwidth, the number of wave trains per second, and PSD. The extent of differences between the group of the first stage Parkinson's disease patients and the healthy volunteers in the space of these parameters is investigated. The ROC-analysis is used for this purpose. The functional dependence of AUC (the area under a ROC-curve) on the boundaries of the ranges of these parameters is analyzed. Using this method, we have identified three frequency ranges, where differences between the group of the patients and the healthy volunteers were discovered. The report describes the results of an investigation of the specificity of these features of early stage of Parkinson's disease.

This research is funded by the program of Presidium of RAS No. I.33II.

- [1] *Sushkova O., Morozov A., Gabova A.* Data Mining in EEG Wave Trains in Early Stages of Parkinson's Disease // Advances in Soft Computing. LNCS. — Cham: Springer, 2017. — P.403–412. https://link.springer.com/chapter/10.1007/978-3-319-62428-0_32.

Реконструкция функциональной структуры сложных систем по многоканальным временным рядам

<i>Устинин Михаил Николаевич</i> ^{1*}	ustinin@impb.ru
<i>Рыкунов Станислав Дмитриевич</i> ¹	rykunov@psn.ru
<i>Бойко Анна Ивановна</i> ¹	a.boyko@list.ru
<i>Сычев Вячеслав Викторович</i> ¹	sychyov@yahoo.com
<i>Молчанова Дина Альбертовна</i> ¹	dm@pochtamt.ru

¹Пушино, Институт математических проблем биологии РАН

Предложен новый метод для реконструкции функциональной структуры сложных систем по излучаемым ими полям. Во время эксперимента массив измерителей, распределенных в пространстве, регистрирует магнитное или электрическое поле в течение некоторого времени. Преобразование Фурье полного временного ряда приводит к детальному спектру, содержащему десятки тысяч частотных компонент. Для каждой частоты восстанавливается многоканальный временной ряд, методом анализа независимых компонент выделяются когерентные составляющие, каждая из которых имеет постоянный паттерн поля. Массовое решение обратной задачи для данного поля и заданной модели источников дает пространственное распределение источников всех компонент — функциональную томограмму. Метод использовался для восстановления функциональной структуры мозга, сердца и скелетных мышц. Также он был успешно применен для локализации магнитных наночастиц без намагничивания и механического смещения образца. Полученные результаты разумно интерпретируются физически и анатомически, что позволяет изучать внутреннюю функциональную структуру различных сложных систем.

Работа поддержана грантами РФФИ № 16-07-00937, № 16-07-01000, № 17-07-00677, № 17-07-00686, № 17-29-02178, программой фундаментальных исследований Президиума РАН № I.33П.

- [1] Polikarpov M. A., Ustinin M. N., Rykunov S. D., Yurenaya A. Y., Naurzakov S. P., Grebenkin A. P., Panchenko V. Y. 3D imaging of magnetic particles using the 7-channel magnetoencephalography device without pre-magnetization or displacement of the sample // Journal of Magnetism and Magnetic Materials. 2017. Vol. 427. P. 139–143. doi:10.1016/j.jmmm.2016.10.055.

Reconstruction of the functional structure of complex systems from the multichannel time series

*Ustinin Mikhail*¹*

ustinin@impb.ru

*Rykunov Stanislav*¹

rykunov@psn.ru

*Boyko Anna*¹

a.boyko@list.ru

*Sychev Vyacheslav*¹

sychyov@yahoo.com

*Molchanova Dina*¹

dm@pochtamt.ru

¹Pushchino, Institute of Mathematical Problems of Biology RAS

A new method was proposed to reconstruct the functional structure of complex systems from the irradiated fields. During the experiment the set of sensors, distributed in space, registers magnetic or electrical field for some time. The Fourier transform of the long time series produces detailed spectrum, containing tens of thousands of frequency components. For each frequency the multichannel time series are reconstructed, and then coherent components are extracted by the independent component analysis. Each component manifests the constant field pattern. Massive solution of the inverse problem for the given field and source model is giving the functional tomogram – spatial distribution of multiple sources. The method was used to reconstruct the functional structure of the brain, heart and skeletal muscles. Then it was successfully applied to localize magnetic nanoparticles without pre-magnetization or displacement of the sample. The results obtained can be reasonably interpreted physically and anatomically, making it possible to study inner functional structure of various complex systems.

The research was partly supported by the RFBR (grants 16-07-00937, 16-07-01000, 17-07-00677, 17-07-00686, 17-29-02178) and by the Program of the Presidium of RAS I.33P.

- [1] *Polikarpov M. A., Ustinin M. N., Rykunov S. D., Yurenja A. Y., Nazarkov S. P., Grebenkin A. P., Panchenko V. Y.* 3D imaging of magnetic particles using the 7-channel magnetoencephalography device without pre-magnetization or displacement of the sample // *Journal of Magnetism and Magnetic Materials*. 2017. Vol. 427. P. 139–143. doi:10.1016/j.jmmm.2016.10.055.

Повышение достоверности молекулярных ДНК-вычислений: решение обратной задачи спектроскопии комбинационного рассеяния с помощью искусственных нейронных сетей

Доленко Татьяна Альдефонсовна^{1,2,3} tdolenko@mail.ru
Буриков Сергей Алексеевич^{1,2} sergey.burikov@gmail.com
*Вервальд Екатерина Николаевна*¹ kl128@mail.ru
Ефиторов Александр Олегович^{2,3*} a.efitorov@sinp.msu.ru
Лаптинский Кирилл Андреевич^{1,2} laptinskij@physics.msu.ru
*Сарманова Ольга Эдуардовна*¹ helga-sharman@rambler.ru
*Доленко Сергей Анатольевич*² dolenko@srd.sinp.msu.ru

¹Москва, МГУ им. М. В. Ломоносова, Физический факультет

²Москва, МГУ им. М. В. Ломоносова, НИИЯФ им. Д. В. Скобельцына

³Москва, НИЯУ МИФИ

Создание биокомпьютеров требует повышения достоверности результатов молекулярных вычислений. В качестве экспрессного бесконтактного метода определения типа и концентрации каждого из четырех азотистых оснований (АО) ДНК, участвующих в процессе вычислений, авторы предлагают использовать спектроскопию комбинационного рассеяния света.

Наличие в спектре линий-маркеров (ЛМ) позволяет определить присутствие АО, а построение калибровочной прямой – его концентрацию. Проблемы – перекрывание ЛМ разных АО и нелинейная зависимость интенсивностей ЛМ от концентраций АО.

Для преодоления данных проблем авторы использовали искусственные нейронные сети (ИНС). Для их обучения были получены спектры многокомпонентных растворов АО с известными концентрациями. Была проведена работа по отбору существенных входных признаков и выбору лучшей топологии ИНС.

Полученная в данной работе погрешность определения потерь «рабочего вещества» молекулярных ДНК вычислений имеет такой же порядок, что и максимально допустимые потери для правильной работы вычислительного биоустройства.

Исследование выполнено за счет гранта Российского научного фонда (проект № 14-11-00579).

- [1] *Dolenko T. A. et al.* Improvement of reliability of molecular DNA computing: solution of inverse problem of Raman spectroscopy using artificial neural networks // *Laser Physics*, V.27, No.2, pp. 025203-1-025203-8.

Improvement of reliability of molecular DNA computing: solution of inverse problem of Raman spectroscopy using artificial neural networks

Dolenko Tatiana^{1,2,3} tdolenko@mail.ru
Burikov Sergey^{1,2} sergey.burikov@gmail.com
*Vervald Ekaterina*¹ kl128@mail.ru
Efitorov Alexander^{2,3}★ a.efitorov@sinp.msu.ru
Laptinskiy Kirill^{1,2} laptinskiy@physics.msu.ru
*Sarmanova Olga*¹ helga-sharman@rambler.ru
*Dolenko Sergey*² dolenko@srd.sinp.msu.ru

¹Moscow, Lomonosov MSU, Department of Physics

²Moscow, Lomonosov MSU, Skobeltsyn Institute of Nuclear Physics

³Moscow, National Research Nuclear University MEPhI

One of the main problems of creating biocomputers is improvement of reliability of molecular computing. The authors suggest using Raman spectroscopy as an express non-contact method to determine the type and concentration of each of the four nitrogenous bases (NB) of deoxyribonucleic acid (DNA) involved in the computation process.

Existence of spectral marker lines (ML) allows detecting a NB, and creating calibration dependencies allows one to determine its concentration. A problem of this approach is its applicability only for single-component solutions, due to overlapping of multiple ML of different NBs. Another problem is the non-linear dependence of the ML intensities on the concentrations of NBs required for computing.

To solve these problems, the authors used artificial neural networks (ANN). A special dataset of spectra of multi-component NB solutions with known concentrations was obtained and used to train the ANN. A special investigation was carried out to select the significant input features and the best topology of the ANN.

The accuracy of determining the loss of the "working substance" of molecular DNA computation obtained in this study has the same order of magnitude as the maximum allowable loss for proper operation of a computing biostructure.

This study has been performed at the expense of the Russian Science Foundation, grant 14-11-00579.

- [1] *Dolenko T. A. et al.* Improvement of reliability of molecular DNA computing: solution of inverse problem of Raman spectroscopy using artificial neural networks // *Laser Physics*, V.27, No.2, pp. 025203-1-025203-8.

Спектрально-аналитический метод распознавания неточных повторов в символьных последовательностях

Панкратов Антон Николаевич

pan@impb.ru

Пушино, ИМПБ РАН – филиал ИПМ им. М. В. Келдыша РАН

Предложены теоретическое обоснование и алгоритмическая реализация спектрально-аналитического метода распознавания повторов в символьных последовательностях. Теоретическое обоснование основывается на теореме об эквивалентном представлении символьной последовательности вектором непрерывных характеристических функций. Сравнение фрагментов характеристических функций производится в стандартной метрике в евклидовом пространстве коэффициентов разложения рядов Фурье по ортогональным многочленам. Существенным свойством данного подхода является способность оценивать повторы на разных масштабах. Другим важным свойством метода является возможность эффективного распараллеливания по данным. При разработке алгоритмов предпочиталась схема вычислений с минимальным количеством обращений к оперативной памяти, подразумевающая повторяющиеся и отложенные вычисления. В данной парадигме разработан алгоритм вычисления коэффициентов разложения по ортогональным многочленам за счет использования рекуррентных соотношений. Показано, что алгоритм вычисления коэффициентов разложения по ортогональным многочленам может быть эффективно векторизован за счет вычисления с фиксированной длиной вектора. Распараллеливание и векторизация реализованы с использованием стандарта OpenMP и расширения Cilk Plus языка C/C++.

Работа поддержана грантом РФФИ, грант 15-29-07063.

- [1] *Панкратов А. Н., Тетуев Р. К., Пятков М. И., Тойгильдин В. П., Попова Н. Н.* Спектрально-аналитический метод распознавания неточных повторов в символьных последовательностях // Труды Института системного программирования РАН Том 27. Выпуск 6. 2015 г. Стр. 335-344. http://www.ispras.ru/proceedings/isp_27_2015_6/isp_27_2015_6_335/.

Spectral Analytical Method of Recognition of Inexact Repeats in Character Sequences

Pankratov Anton

pan@impb.ru

Pushchino, IMPB RAS - Branch of KIAM RAS

Proposed are theoretical basis and algorithmic implementation of spectral-analytical method of recognition of repeats in character sequences. The theoretical justification is based on the theorem on equivalent representation of the character sequence by the vector of continuous characteristic functions. Comparison of fragments of characteristic functions is performed in the standard metric in Euclidean space of expansion coefficients of the Fourier series of orthogonal polynomials. An essential feature of this approach is the ability to evaluate repeats at different scales. Another important feature is the possibility of efficient parallelization of data. In the development of algorithms we preferred scheme of computing with a minimal amount of references to memory, implying repetitive calculations and evaluations on demand. In this paradigm, proposed is an algorithm for calculating the coefficients of expansions in the orthogonal polynomials through the use of recurrence relations. It is shown that the algorithm for calculating the coefficients of expansions in the orthogonal polynomials can be effectively vectorized by computing with a fixed vector length. Parallelization and vectorization implemented using the OpenMP standard and extension Cilk Plus of language C/C++. The developed method effectively scales, depending on the parameters of the problem and the number of processor cores on systems with shared memory.

This research is funded by RFBR, grant 15-29-07063.

- [1] *Pankratov A. N., Tetuev R. K., Pyatkov M. I., Toigildin V. P., Popova N. N.* Spectral Analytical Method of Recognition of Inexact Repeats in Character Sequences // Proceedings of the Institute for System Programming, vol. 27, issue 6, 2015, pp. 335-344. http://www.ispras.ru/en/proceedings/isp_27_2015_6/isp_27_2015_6_335/.

Параллельный алгоритм глобального выравнивания протяжённых аминокислотных и нуклеотидных последовательностей

Тетюев Руслан Курманбиевич^{1*} ruslan.tetuev@gmail.com

*Пятков Максим Иванович*¹ mpyatkov@gmail.com

*Панкратов Антон Николаевич*¹ pan@impb.ru

¹Пушкино, ИМПБ РАН – филиал ИПМ им. М. В. Келдыша РАН

Разработан параллельный алгоритм для глобального выравнивания протяжённых последовательностей. Алгоритм использует произвольную матрицу сходства, аффинную систему штрафов за разрыв и систему конечных штрафов, заданные отдельно для каждой последовательности. Параметрами параллельного алгоритма являются шаги сетки, которая разбивает матрицу глобального выравнивания на блоки. Найдены оптимальные значения этих параметров как по памяти, так и по быстродействию. Показано, что при оптимальном выборе параметров по памяти алгоритм позволяет выравнивать протяжённые последовательности длины L , используя объём памяти $O(L^{4/3})$. Также показано, что алгоритм идеально масштабируется на многоядерных системах, демонстрируя суперлинейное ускорение.

Алгоритм реализован на языке JavaScript в виде высокопроизводительного параллельного веб-приложения, загружаемого с сервера и выполняемого на стороне клиента в виртуальной машине веб-браузера. Алгоритм со значениями параметров по умолчанию выдаёт результаты соответствующие алгоритму, используемому в сервисе глобального выравнивания NCBI BLAST, но превосходит его как по эффективности работы, так и по длине сравниваемых последовательностей, позволяя выравнивать последовательности до 10^6 нуклеотидов.

Работа поддержана грантом РФФИ, грант 15-29-07063.

- [1] *Тетюев Р. К., Пятков М. И., Панкратов А. Н.* Параллельный алгоритм глобального выравнивания протяжённых аминокислотных и нуклеотидных последовательностей // Математическая биология и биоинформатика, 2017. Т. 12. № 1. С. 137–150. <http://www.matbio.org/article.php?id=306>.

Parallel algorithm for global alignment of long aminoacid and nucleotide sequences

Tetuev Ruslan^{1*}

ruslan.tetuev@gmail.com

*Pyatkov Maxim*¹

mpyatkov@gmail.com

*Pankratov Anton*¹

pan@impb.ru

¹Pushchino, IMPB RAS - Branch of KIAM RAS

A parallel algorithm is developed for the global alignment of long DNA and protein sequences. The algorithm uses an arbitrary substitution matrix. An affine internal and end gap penalty systems might be set up separately for each of the two sequences. The possibility to control the choice of an optimal alignment from the set of alternative ones is implemented. The two parameters of the parallel algorithm which are called grid steps were introduced. They are used to split the global alignment matrix into blocks of specified size. The analysis of these parameters was performed in order to obtain optimal values that reduce either time complexity or memory complexity of the algorithm. It is shown that when using the block sizes that provide lowest memory complexity, the algorithm allows to align two long sequences of length L using the memory of size $O(L^{4/3})$. It is shown that the algorithm is well scaled on multi-core systems, reaching superlinear execution time acceleration.

The algorithm is implemented in the JavaScript language as a high-performance parallel web application that is downloaded from the server and executed on the client side in the virtual machine of the web browser. The algorithm with the default parameter values produces the results corresponding to the algorithm used in the NCBI BLAST global alignment service, but it exceeds it both in efficiency and in the length of the sequences compared, allowing the alignment of sequences up to 10^6 nucleotides.

This research is funded by RFBR, grant 15-29-07063.

- [1] *Tetuev R., Pyatkov M., Pankratov A.* Parallel algorithm for global alignment of long aminoacid and nucleotide sequences // *Math. Biol. Bioinf.* 2017. 12(1): 137-150. <http://www.matbio.org/article.php?lang=eng&id=306>.

Исследование межспиральных углов в структурных мотивах, образованных двумя спиральями

*Тихонов Дмитрий Анатольевич*¹ dmitry.tikhonov@gmail.com

Куликова Людмила Ивановна^{1*} likulikova@mail.ru

*Ефимов Александр Васильевич*² efimov@protres.ru

¹Пушино, ИМПБ РАН – филиал ИПМ им. М. В. Келдыша РАН

²Пушино, Институт белка РАН

В данной работе проведен анализ распределения межспиральных углов в парах связанных между собой перетяжками спиралей в пространственных структурах белковых молекул. Были разработаны правила отбора спиральных пар в структурах белковых молекул из Protein Data Bank. Полученное множество спиральных пар было проанализировано с целью его классификации и установления закономерностей структурной организации. По критерию пересечения проекций спиралей на параллельные плоскости, проходящие через оси спиралей, полученное множество было разбито на три подмножества. Показано, что распределение всех типов спиральных пар, не имеющих пересечения проекций спиралей, охватывает практически весь диапазон углов с одним максимумом в области прямого угла. Большинство пар этого множества составляют спиральные пары, состоящие из α - и 3_{10} -спиралей, а множества с пересечением только проекций спиралей – спиральные пары, образованные двумя α -спиральями. Также показано, что образованные двумя α -спиральями спиральные пары составляют абсолютное большинство пар множества с пересечением проекций и осей спиралей. При этом значительная часть указанных пар имеет острый угол ($20^\circ \leq \varphi \leq 60^\circ$) между осями спиралей. Проведен анализ распределения всех типов спиральных пар, принадлежащих различным множествам, в зависимости от длины перетяжки. Показано, что во всех множествах больше всего исследуемых структур с короткой перетяжкой.

Работа поддержана грантом РФФИ № 16-01-00692.

- [1] *Тихонов Д. А., Куликова Л. И., Ефимов А. В.* Исследование межспиральных углов в структурных мотивах, образованных двумя спиральями // Математическая биология и биоинформатика, 2017. — Т. 12. — № 1. — С. 83–101. doi: 10.17537/2017.12.83.

The study of interhelical angles in the structural motifs formed by two helices

*Tikhonov Dmitrii*¹

dmitry.tikhonov@gmail.com

Kulikova Liudmila^{1*}

likulikova@mail.ru

*Efimov Aleksandr*²

efimov@protres.ru

¹Pushchino, IMPB RAS - Branch of KIAM RAS

²Pushchino, Institute of Protein Research RAS

In this paper a statistical analysis of distributions of inter-helical angles in pairs of consecutive and connected α -helices in spatial structures of proteins is presented. A number of rules for selection of the helical pairs from a set of protein structures obtained from the Protein Data Bank (PDB) were developed. The set of helical pairs has been analyzed for the purpose of classification and finding out the features of protein structural organization. All pairs of connected helices were divided into three subsets according to the criterion of crossing of projections of the helices on parallel planes, which pass through the axes of the helices. It is shown that the distribution of all types of helical pairs, whose projections do not cross each others, covers almost the entire range of inter-helical angles. The distribution have a single maximum which is close to right angle. Most pairs in this set constitute helical pairs consisting of α - and 3_{10} -helices, and most pairs with the crossing projections of helices are helical pairs formed by two α -helices. It is also shown that a great amount of the pairs of connected α -helices has acute angle ($20^\circ \leq \varphi \leq 60^\circ$) between the axes of the helices. The distribution of all types of helical pairs depending on the length of the inter-helical connections was also analyzed. It is shown that the structures with short connections occur most often in all the subsets.

This work was supported by RFBR, grant 16-01-00692.

- [1] *Tikhonov D., Kulikova L., Efimov A.* The study of interhelical angles in the structural motifs formed by two helices // *Mathematical Biology and Bioinformatics*, 2017. — V. 12. — No.1. — p. 83–101. doi: 10.17537/2017.12.83.

Оценка пластического ресурса каркаса критически важного объекта при воздействии максимального расчетного землетрясения

Горнов Александр Юрьевич^{1*} gornov@icc.ru
*Бержинский Юрий Анатольевич*² berj@crust.irk.ru
*Финкельштейн Евгения Александровна*¹ finkel@icc.ru
*Зароднюк Татьяна Сергеевна*¹ tz@icc.ru
*Ордынская Алиса Павловна*² berj@crust.irk.ru

¹Иркутск, Институт динамики систем и теории управления имени В. М. Матросова СО РАН

²Иркутск, Институт земной коры СО РАН

Предложенная в работе модификация математической модели сложного упругопластического деформирования механической системы описывает поведение каркаса сооружения в упругой, упругопластической и пластической стадиях работы. Взаимный переход между стадиями происходит при достижении проекциями реакций системы значений, находящихся на поверхности текучести, задаваемой в виде эллипса. Интенсивность максимального расчетного землетрясения принята равной 8 баллам по шкале MSK-64 с повторяемостью один раз в 5000 лет, что соответствует карте Общего сейсмического районирования ОСР-15С. При моделировании пластических деформаций каркаса под воздействием на него сейсмических нагрузок учитывается эффект упруговязкого затухания по Релею.

Реализована методика расчета пластического ресурса каркаса, включающая модуль загрузки данных инструментальных акселерограмм. Проведенные вычислительные эксперименты позволили получить данные о поведении сооружения при воздействиях землетрясений с различным спектральным составом и сделать предварительный деформационный анализ модельного объекта.

Работа поддержана грантом РФФИ № 17-47-380008.

- [1] *Бержинский Ю. А., Горнов А. Ю., Финкельштейн Е. А., Ордынская А. П.* Исследование сложного упругопластического деформирования механических систем с несколькими степенями свободы с помощью вычислительного эксперимента // Информационные и математические технологии в науке и управлении, Иркутск: Изд-во ИСЭМ СО РАН, No 2. — С. 62–70.

The estimation for the plastic resource of the framework of a critically important object under the action of the maximum design earthquake

Gornov Alexander^{1*}

gornov@icc.ru

*Berzhinsky Yuri*²

berj@crust.irk.ru

*Finkelstein Evgeniya*¹

finkel@icc.ru

*Zarodnyuk Tatiana*¹

tz@icc.ru

*Ordynskaya Alisa*²

berj@crust.irk.ru

¹Irkutsk, Matrosov Institute for System Dynamics and Control Theory
SB RAS

²Irkutsk, Institute of the Earth's crust SB RAS

The proposed modification of the mathematical model of the complex elastic-plastic deformation of a mechanical system describes the behavior of the framework in the elastic, elastic-plastic and plastic work stages. The reciprocal transition between stages occurs when the projections of the system's reactions reach the values on the fluidity surface specified in the form of an ellipse. The intensity of the maximum calculated earthquake is taken equal to 8 points on the MSK-64 scale with a frequency of once every 5,000 years, which corresponds to the map of the General Seismic Zoning of OCP-15C. When modeling plastic deformations of the framework under the influence of seismic loads on it, the effect of elastic-viscous Raymond damping is taken into account.

A technique for calculating the plastic framework resource, including a data loading module for instrumental accelerograms, was implemented. The computational experiments carried out made it possible to obtain data on the structure behavior under the influence of earthquakes with different spectral composition, and to make a preliminary deformation analysis of the model object.

This research is funded by RFBR, grant 17-47-380008.

- [1] *Berzhinsky Yu., Gornov A., Finkelstein E., Ordynskaya A.* The program for calculation of complex elastic-plastic deformation of mechanical systems with multiple degrees of freedom // Information and mathematical technologies in science and management, Irkutsk: ISEM SB RAS, No 2. — P. 62–70.

Моделирование рельефа на поверхности сферы, эллипсоида вращения и трехосного эллипсоида

Флоринский Игорь Васильевич

iflor@mail.ru

Пушино, ИМПБ РАН – филиал ИПМ им. М. В. Келдыша РАН

Моделирование рельефа (геоморфометрия) широко используется при решении различных разномасштабных задач в науках о Земле и планетных исследованиях. Существующий алгоритмический аппарат геоморфометрии может корректно применяться для работы с моделями рельефа, заданными как на плоских квадратных сетках, так и на сетках сфероидических трапеций на поверхности сферы и эллипсоида вращения [1].

При моделировании рельефа Земли, Марса, Луны, Венеры и Меркурия расчеты на сетках сфероидических трапеций тривиальны. Это связано с тем, что форма указанных небесных тел описывается эллипсоидом вращения или сферой. Для этих поверхностей существуют хорошо разработанная теория и вычислительные алгоритмы решения главных геодезических задач.

Для описания формы малых небесных тел во многих случаях целесообразно применять трехосный эллипсоид. Однако алгоритмический аппарат, предназначенный для моделирования рельефа на поверхности трехосного эллипсоида, не разработан, так как главные геодезические задачи для случая трехосного эллипсоида решены только в самом общем виде.

Показано, что, если модель рельефа небесного тела или его фрагмента задана на сетке сфероидических трапеций с использованием системы геодезических координат трехосного эллипсоида, то для морфометрических расчетов необходимо перейти в систему эллиптических координат и способом Якоби определить линейные размеры элементов скользящего сфероидического трапецеидального окна. Альтернативным вариантом может служить способ Красовского для определения длин дуг параллелей и меридианов трехосного эллипсоида.

Работа поддержана грантом РФФИ № 15-07-02484.

- [1] *Florinsky I. V.* Spheroidal equal angular DEMs: the specificity of morphometric treatment // Transactions in GIS, 2017, Vol. 21. — doi:10.1111/tgis.12269.

Terrain modelling on a surface of a sphere, ellipsoid of revolution, and triaxial ellipsoid

Florinsky Igor

iflor@mail.ru

Pushchino, IMPB RAS - Branch of KIAM RAS

Terrain modelling (geomorphometry) is widely used to solve various multiscale problems of the Earth and planetary sciences. Existing algorithms of geomorphometry can be applied to terrain models given with plane square grids or spheroidal equal angular grids on a surface of an ellipsoid of revolution or a sphere [1].

Computations on spheroidal equal angular grids are trivial for terrain modelling of the Earth, Mars, the Moon, Venus, and Mercury. This is because forms of the mentioned celestial bodies can be described by an ellipsoid of revolution or a sphere. For these surfaces, there are well-developed theory and computational algorithms to solve main geodetic problems.

It is advisable to apply a triaxial ellipsoid for describing forms of small moons and asteroids. However, there are no terrain modelling algorithms intended for a surface of a triaxial ellipsoid. This is because for the case of a triaxial ellipsoid, main geodetic problems are solved in very general form.

It is demonstrated that if a terrain model of a celestial body or its portion is given with a spheroidal equal angular grid using geodetic coordinate systems of a triaxial ellipsoid, to derive models of local morphometric variables, one should turn to the elliptical coordinate system and then determine linear sizes of spheroidal trapezoidal moving window elements by the Jacobi solution. Alternatively, one can use the Krasovsky method for determination of the lengths of arcs of parallels and meridians of a triaxial ellipsoid.

This research is funded by RFBR, grant 15-07-02484.

- [1] *Florinsky I. V.* Spheroidal equal angular DEMs: the specificity of morphometric treatment // Transactions in GIS, 2017, Vol. 21. — doi:10.1111/tgis.12269.

Разработка алгоритма контроля перемещения изделий между зонами складирования на основе статистического анализа временных рядов данных с их идентификаторов для систем промышленной автоматизации

Астафьев Александр Владимирович^{1*}

Alexandr.Astafiev@mail.ru

*Орлов Алексей Александрович*¹

AlexeyAlexOrlov@gmail.com

¹Муром, Муромский институт (филиал) ВлГУ

В настоящее время, в связи с необходимостью перехода к новым интеллектуальным цифровым производственным технологиям и реализации международных стандартов качества требуется введение новых наукоемких подходов к контролю движения продукции и средств малой механизации складских хозяйств.

Статья посвящена разработке технологии, математического и программного обеспечения для построения систем позиционирования и контроля движения средств малой механизации промышленных предприятий на основе методов радиочастотной идентификации и технического зрения, что будет являться базой для создания высокоэффективных интеллектуальных систем контроля движения продукции на промышленных предприятиях.

В рамках работы были проведены экспериментальные исследования на промышленном предприятии региона. В ходе проведения исследований маркированная металлическая продукция перемещалась между стеллажами средствами малой механизации, в частности мостовым краном с грузозахватной траверсой. Экспериментальные исследования показали корректность работы алгоритма определения текущей зоны складирования в лабораторных и производственных условиях.

Работа поддержана грантом Президента РФ № МК-991.2017.9.

- [1] Orlov A. A., Provotorov A. V., Astafiev A. V. Methods and algorithms of automated two-stage visual recognition of metal-rolling billets // Automation and Remote Control, Pleiades Publishing, Ltd, 2016. — С. 1099–1105. <https://doi.org/10.1134/S000511791606014X>.

Development of an algorithm for controlling the movement of products between storage areas based on the statistical analysis of time series data from their identifiers for industrial automation systems

*Astafiev Alexandr*¹*

Alexandr.Astafiev@mail.ru

*Orlov Alexey*¹

AlexeyAlexOrlov@gmail.com

¹Murom, Murom Institute Branch of VISU

Currently, because of the need to move to new intellectual digital production technologies and implement international quality standards, it is necessary to introduce new science-based approaches to controlling the movement of products and small-scale mechanization of warehouses.

The article is devoted to the development of technology and software for the construction of positioning and control systems for small mechanization in industrial plants based on radio frequency identification methods and technical vision, which will be the basis for creating highly efficient intelligent systems for controlling the product movement in industrial enterprises. Application of these knowledge-intensive technologies makes it possible to automate the processes of controlling the product movement at industrial plants, ultimately, to increase the efficiency and reliability of transportation control and warehouse inventory control of manufactured products.

Experimental research was carried out at the industrial enterprise of the region. During the research, the labeled metal products were moved between the shelves by means of small-scale mechanization, in particular a bridge crane with a load-carrying beam.

The work is executed at financial support of the grant of the President of the Russian Federation MK-991.2017.9.

- [1] *Orlov A. A., Provotorov A. V., Astafiev A. V.* Methods and algorithms of automated two-stage visual recognition of metal-rolling billets // Automation and Remote Control, Pleiades Publishing, Ltd, 2016. — p. 1099–1105. <https://doi.org/10.1134/S000511791606014X>.

Перспективы использования кодов с коррекцией ошибок в стеганографических системах, использующих в качестве контейнеров аудио файлы

Жарких Александр Александрович^{1*} zharkikh090107@mail.ru
*Червяков Иван Олегович*¹ ivan753.cr@yandex.ru

¹Мурманск, Мурманский государственный технический университет

Цель данного сообщения – формулировка основных принципов применения кодов с коррекцией ошибок в стегосистемах, а также указание перспектив такого применения в случае, когда контейнер является аудиофайлом.

Наиболее известные классы кодов с коррекцией ошибок – классы свёрточных и блочных кодов. Свёрточные коды используются для кодирования больших объёмов информации без разделения её на отдельные блоки. Блочные коды, наоборот, используются после предварительного разбиения информации на отдельные блоки. Среди блочных кодов наиболее эффективными с точки зрения реализации являются циклические коды.

Сформулируем два основных принципа применения кодов с коррекцией ошибок в стегосистеме. Первый связан с тем, что стегосистема строится на основе легальной системы передачи информации. Поэтому необходимо обеспечить качественную передачу контейнера, либо стегоконтейнера в открытом легальном канале передачи. Второй же связан непосредственно с построением стегосистемы. Он сводится к требованию обеспечения высококачественной передачи сообщения в скрытом канале связи.

В случае, когда исходный контейнер является аудиофайлом, информация естественным образом делится на блоки. Так как отсчёты аудиосигнала в аудиофайле записываются одинаковым количеством бит, то отдельный отсчёт, как и объединение нескольких, можно считать отдельным блоком. В силу эффективности реализации циклических кодов перспективным является их применение в обоих каналах стегосистемы.

- [1] *Жарких А. А. Червяков И. О.* Принципы использования кодов с коррекцией ошибок в стеганографических системах, использующих в качестве контейнеров аудио файлы // Машинное обучение и анализ данных, 2017.

Prospects for the use of error correction codes in steganographic systems using audio files as containers

Zharkikh Alexander¹*

zharkikh090107@mail.ru

Chervyakov Ivan¹

ivan753.cr@yandex.ru

¹Murmansk, Murmansk State Technical University

The purpose of this report is to formulate the basic usage principles of error correcting codes in steganographic systems, and also to indicate the prospects for such application in the case when the container is an audio file.

The most well-known classes of codes with error correction are the classes of convolutional and block codes. Convolutional codes are used for encoding large amounts of data without splitting it into separate blocks. Block codes, on the contrary, are used after the preliminary partitioning of information into separate blocks. Among the block codes, cyclic codes are the most effective from the point of view of implementation.

Let's formulate two basic application principles of codes with error correction in the steganographic system. The first principle stems from the fact that the stegosystem is built on the basis of a legal information transfer system. Therefore, it is necessary to ensure the quality transfer of the container (if there is no embedding) or stegocontainer (if there is integration) in the open legal transmission channel. The second principle is directly related to the construction of stegosystem. It boils down to the requirement for high-quality message transmission in a hidden communication channel.

In the case where the source container is an audio file, the information is naturally divided into blocks. Since audio samples in an audio file are recorded with the same number of bits, a separate sample can be considered as a separate block. Blocks can also be obtained by combining multiple samples of an audio signal. In view of the realisation efficiency of cyclic codes, its usage in both channels of the stegosystem is perspective.

- [1] Zharkikh A. A. Chervyakov I. O. The usage principles of error correcting codes in steganographic systems using audio files as containers // Journal of Machine Learning and Data Analysis, 2017.

Применение гауссовских вращений для защиты графической информации

*Каркищенко Александр Николаевич*¹ karkishalex@gmail.com
Мнухин Валерий Борисович^{1*} mnukhin.valeriy@mail.ru

¹Таганрог, Южный федеральный университет

В работе рассматриваются цифровые изображения над «конечными комплексными полями». Вводится преобразование *гауссова вращения* таких изображений и доказываемся, что при определенных условиях его результат напоминает несколько уменьшенных и повернутых копий оригинала, несмотря на то, что эти «копии» образованы различными пикселями. Рассмотрена возможность создания на основе гауссовых вращений защитных фонов и текстур для предотвращения несанкционированного изменения документов. Приведен метод верификации защищенных таким образом документов.

Работа поддержана грантами РФФИ № 16-07-00648-а и № 17-20-02017-офи_м_РЖД.

- [1] *Каркищенко А. Н., Мнухин В. Б.* Применение гауссовских вращений для защиты графической информации // Машинное обучение и анализ данных, 2017.

Gaussian rotations for graphic information protection

*Karkishchenko Alexander*¹

karkishalex@gmail.com

Mnukhin Valeriy^{1*}

mnukhin.valeriy@mail.ru

¹Taganrog, Southern Federal University

Digital images over “finite complex planes” are considered jointly with transformations of *gaussian rotations*. It is proved that under some special conditions results of such transformations seem to be formed by several zoomed out copies of the rotated original, though all such “copies” are formed by different pixels of the original image. Based on gaussian rotations some methods for tamper resistant protection of graphic information are considered. A method for verification of protected information is also introduced.

This research is funded by RFBR, grants 16-07-00648-a and 17-20-02017-ofi_m_RZhD.

- [1] *Karkishchenko A., Mnukhin V.* Gaussian rotations for graphic information protection // *Journal of Machine Learning and Data Analysis*, 2017.

Методы построения систем анализа состояния распределённых информационных систем на основе беспризнакового распознавания образов

*Руднев Дмитрий Олегович*¹
Сычугов Алексей Алексеевич^{1*}

dima_rudnev1@mail.ru
xru2003@list.ru

¹Тула, Тульский государственный университет

В настоящее время выделяют два основных подхода к анализу состояния распределённых информационных систем (РИС): централизованный и распределённый. Централизованный подход заключается в сборе информации о работе всех элементов РИС и последующего анализа всего объёма данных в одном элементе РИС. К недостаткам такого подхода относятся высокие требования к вычислительным ресурсам элемента РИС. Распределённый подход заключается в том, что каждый элемент сам анализирует свое состояние на основе некоторого ранее определённого решающего правила. Недостатком такого подхода, является низкая адаптивность.

Предложен гибридный подход к анализу состояния РИС, позволяющий с одной стороны обеспечить высокую адаптивность, а с другой, обеспечить оптимальное использование вычислительных ресурсов.

Разработанный подход основан на том, что каждый элемент передает на другой узел не непосредственно свои признаки, а значения меры схожести своих признаков на множество заранее заданных других элементов РИС.

Проведенный эксперимент по определению типа активности пользователя мобильного устройства показал, что гибридный подход имеет точность определения состояния выше, чем при использовании распределённого подхода и при этом гибридный подход требует меньше вычислительных ресурсов чем централизованный подход.

Работа поддержана грантом РФФИ № 16-07-01008.

- [1] *Руднев Д. О., Сычугов А. А.* Методы построения систем анализа состояния распределённых информационных систем на основе беспризнакового распознавания образов // Известия ТулГУ, Тула: Технические науки, Выпуск 10, 2017.

Methods for constructing systems for analyzing the state of distributed information systems based on unrecognized pattern recognition

*Rudnev Dmitry*¹

dima_rudnev1@mail.ru

Sychugov Alexey^{1*}

xru2003@list.ru

¹Tula, Tula State University

Currently, there are two main approaches to analyzing the state of distributed information systems (DIS): centralized and distributed. The centralized approach is to collect information on the operation of all DIS elements and the subsequent analysis of the entire volume of data in one element of DIS. The disadvantages of this approach include high requirements for the computational resources of the DIS element. A distributed approach is that each element itself analyzes its state on the basis of some previously defined decision rule. The disadvantage of this approach is low adaptability.

A hybrid approach to analyzing the state of DIS is proposed, which, on the one hand, allows for high adaptivity, and on the other hand, ensures optimal use of computing resources.

The developed approach is based on the fact that each element transmits to its other node not directly its characteristics, but the values of the measure of the similarity of its characteristics to the set of predetermined other elements of the RIS.

The conducted experiment to determine the type of activity of the user of the mobile device showed that the hybrid approach has an accuracy of determining the state higher than using a distributed approach, and the hybrid approach requires less computational resources than a centralized approach.

This research is funded by RFBR, grant 16-07-01008.

- [1] *Rudnev D., Sychugov A.* Methods for constructing systems for analyzing the state of distributed information systems based on unrecognized pattern recognition // *Izvestiya of Tula State University, Technical Sciences.* Issue 10, 2017

Обнаружение вредоносного программного обеспечения с использованием иммунных детекторов

*Токарев Вячеслав Леонидович*¹

tokarev22@yandex.ru

Сычугов Алексей Алексеевич^{1*}

xru2003@list.ru

¹Тула, Тульский государственный университет

Вредоносное программное обеспечение предназначено для повреждения сети или компьютерных ресурсов. Сегодня бум появления вредоносных программ заставляет исследователей разрабатывать новые методы для защиты компьютеров и сетей. Среди трех основных методов, эвристический, сигнатурный и поведенческий, используемых для обнаружения вредоносных программ, наиболее распространенным является эвристический метод.

Предлагается вариант эвристического метода обнаружения вредоносных программ, основанный на использовании иммунных детекторов в виде скрытой Марковской модели (СММ).

Рассмотрена общая схема функционирования искусственных иммунных систем.

Для вычисления апостериорных вероятностей последовательности скрытых состояний при наличии последовательности наблюдений в СММ применен алгоритм «прямого-обратного» хода, а для нахождения неизвестных параметров СММ использован алгоритм Баума-Велша.

Описан жизненный цикл иммунных детекторов, включающий в себя этапы генерации, обучения, отбора и функционирования, состоящий из активизации иммунных детекторов и формирования иммунной памяти. Предложен общий алгоритм функционирования системы обнаружения вредоносного программного обеспечения.

Экспериментальные исследования показали высокую вероятность (в среднем 91%) обнаружения метаморфических вирусов. Предложенный метод может служить основой для построения эффективных систем обнаружения вредоносных программ.

Работа поддержана грантом РФФИ № 16-07-01008

- [1] *Токарев В. Л., Сычугов А. А.* Обнаружение вредоносного программного обеспечения с использованием иммунных детекторов // Известия ТулГУ, Тула: Технические науки, Выпуск 10, 2017.

Detection of malicious software using immune detectors

*Tokarev Vjacheslav*¹

tokarev22@yandex.ru

Sychugov Alexey^{2*}

xru2003@list.ru

¹Tula, Tula State University

Malicious software is designed to damage network or computer resources. Today, the boom in the appearance of malware forces researchers to develop new methods for protecting computers and networks. Among the three main methods, heuristic, signature and behavioral, used to detect malicious programs, the most common is the heuristic method.

A variant of the heuristic method for detecting malware, based on the use of immune detectors in the form of a Hidden Markov Model (HMM) is proposed.

The general scheme of functioning of artificial immune systems is considered.

To calculate a posteriori probabilities of a sequence of hidden states in the presence of a sequence of observations in HMM, the algorithm of the “forward-reverse” course is applied, and the Baum-Welsh algorithm is used to find the unknown parameters of the HMM.

The life cycle of immune detectors, which includes the stages of generation, training, selection and functioning, consisting of activation of immune detectors and the formation of immune memory is described. A general algorithm for the detection of malicious software is proposed.

Experimental studies have shown a high probability (on average 91%) of detection of metamorphic viruses. Thus, the proposed method can serve as a basis for building effective systems for detecting malicious programs.

This research is funded by RFBR, grant 16-07-01008.

- [1] *Tokarev V., Sychugov A.* Detection of malicious software using immune detectors // *Izvestiya of Tula State University, Technical Sciences.* Issue 10, 2017

Абстракция метода как средства создания понятий

Беликов Александр Николаевич

anbelikov@sfedu.ru

Таганрог, Южный федеральный университет

Разработка программного обеспечения является трудоемким процессом. Как показывают статистические исследования более половины проектов по созданию информационных систем являются не успешными (возникают какие-то трудности, либо вообще не реализовываются). Все существующие подходы основываются на так называемой онтологической парадигме (объектной парадигме), которая описывает понятия (объекты) и связи между ними. Это позволяет создавать отдельные модели определенной предметной области, однако при переходе между предметными областями (разнопредметными знаниями) произвести соответствие понятий (объектов) не представляется возможным. Проблема заключается в переходе между совокупностями элементов различных уровней (например, элементами предметной области и элементами системы) и соответственно между отношениями между ними, и как правило данный переход производится разработчиками ИС эмпирически. В докладе предлагается решение данной проблемы и новый взгляд на проектирования информационных систем в целом. Предлагается рассматривать не понятия (объекта), а способ (средство) их создания. Для этого предлагается методологическая модель описания понятий в виде методов, используемых на разных этапах разработки информационных систем. Данная модель включает такие понятия как метод, методика и норма (набор конкретных действий). Описывается процесс согласования данных понятий, состоящий из двух уровней: вертикальное и горизонтальное согласования.

Работа поддержана грантом РФФИ № 15-07-04250 А.

- [1] *Беликов А. Н., Rogozov Ю. И.* Методологическая модель для согласования этапов разработки информационной систем // Информатизация и связь № 4, Москва, 2017. — с. 61-64.

The abstraction of a method as means of creating concepts

Belikov Alexandr

anbelikov@sfnedu.ru

Taganrog, Southern Federal University

Software development is a time-consuming process. As shown by statistical studies, more than half of the projects for the information systems creation are not successful (there are some difficulties, or not implemented at all). All existing approaches are based on the so-called ontological paradigm (object paradigm), which describes the concepts (objects) and the connections between them. This allows to create separate models of a specific domain, but when you navigate between subject areas (multicomponent knowledge), it is not possible to make the correspondences of concepts (objects). The problem is in the transition between sets of elements of different levels (for example, elements of the domain and elements of the system) and, respectively, between the relations between them, and as a rule this transition is made empirically by IS developers. The report proposes a solution to this problem and a new look at the design of information systems in general. It is suggested to consider the way (means) of objects creation. For this, a methodological model is proposed for describing concepts in the form of methods used at different stages of the information systems development. This model includes such concepts as method, methodology and norm (a set of concrete actions). The process of reconciling these concepts is described, consisting of two levels: vertical and horizontal matching.

The study was supported by RFBR, project 15-07-04250 A.

- [1] *Belikov A., Rogozov Y.* Methodological model for the coordination of the stages of the development of the information systems // *Informationization and Communication* No. 4, Moscow, 2017. — pp.61-64.

Проектирование пользовательского интерфейса ИС на основе семантического подхода

Беликова Светлана Алексеевна

belousova@sfedu.ru

Таганрог, Южный федеральный университет

В настоящее время в области проектирования пользовательских интерфейсов и информационных систем в целом наблюдается ситуация, когда результат разработки не устраивает заказчика. Эта проблема возникает именно потому, что при проектировании интерфейса пользователя не учитывается смысл деятельности пользователя, последовательность и содержание действий, а также теряются требования пользователя. Наблюдается непонимание разработчиками реальных ожиданий потенциальных пользователей системы. В отличие от существующих в данной области подходов, предлагаемый подход к смысловому проектированию учитывает деятельность пользователя в предметной области, за счет чего можно решить проблемы потери смысла предметной области, решить проблему ямы недопонимания между стейкхолдерами процесса проектирования интерфейса, что положительно скажется на эффективности проектируемого интерфейса [1]. Предлагаемый семантический подход заключается в использовании единой знаковой формы взаимодействия пользователя и разработчика интерфейса. Предполагается, что единая знаковая форма способна облегчить процессы анализа предметной области и частично заменить методы, используемые на данном этапе. В качестве единой знаковой формы будет использован механизм действия, который представляет собой таблицу с заранее определенной структурой, повторяющей структуру действия. Содержанием механизма являются составляющие деятельности пользователя, шаблоны технологии построения интерфейса, модель интерфейса, и т.д. в зависимости от уровня декомпозиции механизмов.

Работа поддержана грантом РФФИ № 17-07-00099 А.

- [1] *Belikova S., Rogozov Y., Borisova E.* Semantic approach to information system user interface design // 17th International Multidisciplinary Scientific GeoConference, SGEM 2017, vol. 17, issue 21, Sofia, Bulgaria, 2017. — pp. 673-681.

Information systems user interface design based on semantic approach

Belikova Svetlana

belousova@sfedu.ru

Taganrog, Southern Federal University

Currently in the design of user interfaces and information systems in general, there is a situation where the result of development does not suit the customer. This problem occurs precisely because the designer of the user interface does not take into account the meaning of the user's activities, the sequence and content of the actions, and also the user's requirements are lost during the design process. The developers do not understand the real expectations of potential system users. Unlike the existing approaches in this area, the proposed approach to the semantic design takes into account the user's activity in the subject area, due to which it is possible to solve the problems of the domain meaning loss, to solve the problem of the misunderstanding gap between the stakeholders of the user interface design process. Solving these problems will positively affect the efficiency of the designed interface. To assess this effectiveness, as well as to assess the quality of preserving the meaning of the subject area of the proposed approach, the article proposes to use criteria of relevance and pertinence of the interface [1]. The proposed semantic approach is to use a single sign form of interaction between the user and the developer of the interface. It is assumed that a single sign form can facilitate the analysis of the subject area and partially replace the methods used at this stage. As a single sign form, an action mechanism will be used, which is a table with a predetermined structure that repeats the structure of the action. The content of the mechanism is the components of the user's activity, the templates of the interface construction technology, the interface model, etc. depending on the level of decomposition mechanisms.

The study was supported by RFBR, project 17-07-00099 A.

- [1] *Belikova S., Rogozov Y., Borisova E.* Semantic approach to information system user interface design // 17th International Multidisciplinary Scientific GeoConference, SGEM 2017, vol. 17, issue 21, Sofia, Bulgaria, 2017. — pp. 673-681.

Разработка метода построения процедур вставки данных в структурно-независимые базы данных на основе деятельностного подхода

Грищенко Андрей Сергеевич

agrishenko@sfedu.ru

Таганрог, Южный федеральный университет

Структурно-независимые базы данных широко используются в различных областях для решения определенных задач. Тем не менее, их основной проблемой считается низкая эффективность, которая не позволяет им широко развиваться. Данная работа посвящена проблемам, связанным со вставкой данных. Причинами такой низкой производительности может быть названо применение существующего «объектного» метода построения процедур и, неоднородность формы представления процесса построения процедуры его цели и результату. Этот процесс использует объектную форму представления для процесса построения процедуры вставки данных, в свою очередь, цель и результат представляются действиями. Целью исследования является создание нового метода построения процедур вставки данных в структурно-независимые базы данных однородного действия и основанного на деятельностном подходе. Для осуществления поставленной цели мы введем новое понятие «факт», как минимальную единицу хранения в базе данных, и используем действия на всех этапах построения процедуры вставки данных, используя «механизмы». В результате, используя полученный метод, были построены новые процедуры, которые показали схожие с реляционной БД результаты, которые можно считать удовлетворительными.

Работа поддержана грантом РФФИ № 17-07-00105.

- [1] Rogozov Y., Grisichenko A., Lipko J. Method of creating procedures for inserting data in structure-independent databases based on activity driven approach // Proceedings of 17th International Multidisciplinary Scientific Geoconference “Informatics, Geoinformatics and remote sensing” SGEM-2017, Vol.21, Albena, Bulgaria, 2017. — p. 425–432.

Method of creating procedures for inserting data in structure-independent databases based on activity driven approach

Grishchenko Andrey

agrishenko@sfedu.ru

Taganrog, Southern Federal University

Structure-independent databases are widely used in different spheres for specific tasks solving. Nevertheless, their main problem is a low efficiency that prevents them to be in progress. This work is concerned to the problems associated with data insertion. The causes of this low performance can be called the use of the existing object method of procedures creation and heterogeneity of the data insertion procedures creation process. This process uses object representation form for process of data insertion procedures creation and activity representation form for its purpose and result. Our aim is to create a new method that allows to make insertion procedures creation process homogeneous to the action using activity driven approach. In this way, we implement new concept “fact” as minimal database storage unit and activities in all phases of data insertion procedures creation process using “mechanisms”. As a result, new procedures were built using developed method and they show similar results to relational database procedures, which can be considered satisfactory.

The reported study was partially supported by RFBR, research project 17-07-00105.

- [1] Rogozov Y., Grishchenko A., Lipko J. Method of creating procedures for inserting data in structure-independent databases based on activity driven approach // Proceedings of 17th International Multidisciplinary Scientific Geoconference “Informatics, Geoinformatics and remote sensing” SGEM-2017, Vol.21, Albena, Bulgaria, 2017. — p. 425–432.

Разработка новых методов организации дорожного движения на основе самоорганизации связей

Елькин Дмитрий Максимович

delkin@sfedu.ru

Таганрог, Южный федеральный университет

Управление дорожным движением – одна из наиболее важных и сложных задач современного города. Количество транспортных средств с каждым годом растет быстро развивающимися темпами. Методы дорожного регулирования так же изменяются, но не успевают за ростом транспортного потока. Поэтому требуется разработка других методов и подходов, которые будут решать задачу организации дорожного движения эффективнее чем существующие, а так же менее ресурсозатратно. Были проанализированы следующие известные методы дорожного регулирования: АСУДД, независимое жесткое управление перекрестком; независимое управление перекрестком по информации с детекторов транспорта; фиксированное координированное управление. Все представленные методы оперируют в первую очередь техническими средствами и в этом их недостаток. Нужно начинать организацию дорожного движения на некотором участке дорожной сети с выявления целевой функции движения ее свойств и связей, которые при построении самоорганизуются в процесс движения на основе анализа которого получится выявить необходимые технические средства и правила движения для этого участка. Используя и изменяя свойства мы будем получать различные наборы технических средств и правил движения, которые можно использовать для решения конкретной проблемной ситуации на дорожной сети города. Этот метод позволит разработать решение как для нескольких отдельных проблемных участков дорожной сети, так и для сети города в целом.

- [1] *Елькин Д. М.* Разработка новых методов организации дорожного движения на основе самоорганизации связей // Технологии разработки информационных систем ТРИС-2016 материалы конференции: в 2-х томах, Ростов-на-Дону: Южный Федеральный Университет, 2016. — С. 141–143. https://elibrary.ru/download/elibrary_26639273_50031106.pdf.

Development of new methods for organizing traffic on the basis of self-organization of links

Elkin Dmitry

delkin@sfedu.ru

Taganrog, Southern Federal University

Traffic management is one of the most important and complex tasks of a modern city. The number of vehicles is growing rapidly with each passing year. Methods of road regulation also change, but do not keep up with traffic flow growth. Therefore, it is necessary to develop other methods and approaches that will be solved by a more expensive movement than that, and also less resource-intensive. The following types of road regulation methods were analyzed: ASUDD, independent rigid control of the intersection; independent control of the intersection according to information from transport detectors; fixed coordinated management. All of them are primarily technical means and in this their lack. It is necessary to start the organization of traffic on some part of the road, through which they are organized in the process of movement on the basis of analysis, which will determine the necessary technical means and traffic rules for this site. Using and changing the properties we will receive various sets of hardware and rules that can be used to solve problems of a problem situation on the city's road network. This method allows to solve the solution for several separate problem areas of the road network, and for city networks as a whole.

- [1] *Elkin D.* Development of new methods for organizing traffic on the basis of self-organization of links // Technologies of development of information systems TRIS-2016 conference materials: in 2 volumes, Rostov-on-Don: Southern Federal University, 2016. — p.141–143. https://elibrary.ru/download/elibrary_26639273_50031106.pdf.

NoSQL-подход к хранению данных в конфигурируемой информационной системе

Кучеров Сергей Александрович^{1*}

skucherov@sfedu.ru

*Рогозов Юрий Иванович*¹

yrogozov@sfedu.ru

¹Таганрог, Южный федеральный университет

Переход к информационным системам нового поколения – конфигурируемым, отличающимся как по архитектуре, так и по принципам функционирования обуславливает появление отличных от существующих принципов хранения данных и способов их реализации. Если абстрагироваться от технических и методологических особенностей разработки, можно рассматривать информационную систему как отражение действий пользователя, выполняющихся с применением других средств, но с тем же результатом

По нашему мнению, единственной абстракцией, способной в адекватной предметной области и пользователю форме представить информационную систему является абстракция синтеза действия – фиксированная совокупность характеристик с изменяемым содержанием. Изменение содержания определяет назначение действия, в то время как фиксированная структура представления действия позволяет использовать одну абстракцию на всем протяжении процесса создания системы – от исследования предметной области, до создания и эксплуатации системы. Это, в свою очередь, требует наличия иных методов анализа и синтеза систем, основанных на абстракции действия.

Следовательно, хранилище данных конфигурируемой информационной системы должно быть ориентировано на хранение не отдельных фактов, а действий пользователя, выполнение которых приводит к их получению. В данном докладе представлена модель хранилища конфигурируемой информационной системы, адекватная модели самой системы и концептуальная модель хранилища, описанная с учетом использования NoSQL-технологий.

Работа поддержана грантом РФФИ № 15-07-04033.

- [1] *Kucherov S., Rogozov Yu., Sviridov A.* NoSQL Approach to Data Storing in Configurable Information Systems // Communications in Computer and Information Science, vol. 584, 2016. — p. 120–134.

NoSQL Approach to Data Storing in Configurable Information Systems

Kucherov Sergei^{1*}

skucherov@sfedu.ru

*Rogozov Yury*¹

yrogozov@sfedu.ru

¹Taganrog, Southern Federal University

The transition to information systems of a new generation - configurable, differing both in architecture and in accordance with the principles of operation, causes the appearance of data different from the existing principles and methods of their implementation. If we abstract from the technical and methodological features of the development, we can consider the information system as a reflection of the user's actions performed using other means, but with the same result

In our opinion, the only abstraction capable of presenting an information system in an adequate subject area and the user's form is an abstraction of the synthesis of action—a fixed set of characteristics with a variable content. Changing the content determines the purpose of the action, while the fixed structure of the action view allows one abstraction to be used throughout the system creation process, from subject domain research to the creation and operation of the system. This, in turn, requires the presence of other methods for analyzing and synthesizing systems based on abstraction of action.

Therefore, the data store of a configurable information system should be oriented to storing not individual facts, but user actions, the execution of which leads to their receipt. This report presents a model of a repository of a configurable information system that is adequate to the model of the system itself and the conceptual model of the storage, described taking into account the use of NoSQL technologies.

This research is funded by RFBR, grant 15-07-04033.

- [1] *Kucherov S., Rogozov Yu., Sviridov A.* NoSQL Approach to Data Storing in Configurable Information Systems // Communications in Computer and Information Science, vol. 584, 2016. — p. 120–134.

Разработка механизмов трансляции программных моделей, представленных в формате среды Matlab

Лапшин Вячеслав Сергеевич

vacheslav_93@mail.ru

Таганрог, Инженерно-технологическая академия ЮФУ

В работе рассматривается реализация модуля конвертации программных моделей, представленных в формате среды Matlab и на языке программирования C++. Данный модуль и его дополнительные компоненты разработаны на кафедре вычислительной техники Инженерно-технологической академии Южного федерального университета (ЮФУ).

На сегодняшний день весьма актуальной является проблема организации взаимодействия программных комплексов, в которых исследователи ведут свои разработки по созданию программных моделей различных сложных технических систем. Одним из самых востребованных комплексов является среда разработки Matlab. Также, практически каждая современная среда разработки поддерживает программные модели, реализованные на языке программирования C++. Таким образом, для организации наиболее эффективного процесса разработки программных моделей различных технических систем, разработчикам необходимы инструменты, которые позволят быстро и эффективно преобразовывать самые популярные форматы как отдельных компонент моделей, так и всей модели целиком. В данной работе рассматривается разработка такого инструментария для конвертации моделей на базе программного комплекса «Мультитранслятор». Сам конвертер представляет из себя два взаимодействующих трансляционных модуля. В итоге, произведены эксперименты по конвертации различных программных моделей в ходе которых была произведена оценка работоспособности разработанного модуля.

- [1] *Lapshin V. S., Polenov M. Y.* Development of mechanisms for translating software models presented in the matlab environment format // ITDT2017, Berlin: MVB Marketing- und Verlagsservice des Buchhandels GmbH, 2017. — p. 199–208.

Development of mechanisms for translating software models presented in the matlab environment format

Lapshin Vjacheslav

vacheslav_93@mail.ru

Taganrog, SFedU Engineering and Technology Academy

In this work the implementation of program models' converting module for models in the Matlab format and in C ++ programming language is considered. This module and its additional components are developed at the Department of Computer Engineering in Engineering and Techno-logical Academy of the Southern Federal University (SFedU).

Today, the problem of the interaction organization of the program complexes where re-searchers develop program models of different complex technical systems is very urgent. One of the most popular complex is Matlab development environment. Also, practically each modern development environment supports the program models realized in C ++ language. Thus, for the organization of the most effective development process of program models of different technical systems, the tools are necessary for developers which will allow to transform quickly and effectively the most popular formats of separate components of models, and all model entirely. In this paper the development of such converting tools on the basis of the Multitranslator complex, which is developed earlier at Department of Computer Engineering, is described. The converter consists of two interacting translation modules. This principle of the conversion organization allows to upgrade effectively developed tools adding or changing translation modules. Also, the possible impasses arising during program code translation were analyzed and mechanisms of their overcoming are developed. As a result, experiments on converting of different program models for the assessment of operability of the developed program models converter were made.

- [1] *Lapshin V. Polenov M.* Development of mechanisms for translating software models presented in the matlab environment format // ITDT2017, Berlin: MVB Marketing- und Verlagsservice des Buchhandels GmbH, 2017. — p. 199–208.

Модель информационных потоков для проектирования информационной сети предприятия

Свиридов Александр Славьевич^{1*}

asviridov@sfedu.ru

*Липко Юлия Юрьевна*¹

ylipko@sfedu.ru

¹Таганрог, Южный федеральный университет

Одним из важнейших моментов при проектировании автоматизированных информационных систем (АИС) является рассмотрение информационных потоков, циркулирующих на предприятии. Целью создания модели информационных потоков является как их оптимизация, так и проектирование на ее основе информационной сети предприятия.

Для математического описания информационных потоков подходят матричные модели, которые позволяют исследовать потоки данных между различными элементами структуры в различных разрезах. Модель в таком виде описывается с помощью набора матриц определенного вида, а преобразования модели производятся с помощью стандартных матричных операций. Данный вид модели удобен для обработки на ЭВМ.

Предлагается использовать матричную модель, которая ориентированна на машинную обработку и представлена набором простых матриц, в отличие от известной модели Данилевского Ю.Г., состоящей из единой матрицы сложной структуры. Во-вторых, элементами матриц будут интенсивности, характеризующие параметры документооборота. За счет этого можно значительно сократить количество промежуточных операций.

Построив данную модель можем точно определить интенсивность информационного взаимодействия между всеми подразделениями, объемы передаваемой информации, обоснованно рассчитать характеристики каналов передачи данных. Зная структуру потоков по предприятию и их соотношение внутри каждого из подразделений можно проектировать архитектуру информационной сети, структуру хранилища данных.

Работа поддержана грантом РФФИ № 17-07-00096.

- [1] Свиридов А. С. Метод построения математической модели информационных потоков предприятия // Известия ЮФУ. Технические науки, Таганрог: Издательство ТРТУ, 2004. — С. 152–156.

Model of information flows for the design of enterprise information network

Sviridov Alexander^{1*}

asviridov@sfedu.ru

*Lipko Yulia*¹

ylipko@sfedu.ru

¹Taganrog, Southern Federal University

One of the most important moments in the design of automated information systems (AIS) is the consideration of information flows circulating in the enterprise. The purpose of creating a model of information flows is how to optimize them, and design on its basis the enterprise information network.

The matrix model is suitable for the mathematical description of information flows in an enterprise, which allow one to investigate data flows between different elements of the structure in different dimensions. The model in this form is described using a set of matrices of a certain type, and model transformations are performed using standard matrix operations. Also, this type of model is convenient for presentation and processing on a computer.

It is proposed to use a matrix model that is machine-oriented and is represented by a set of simple matrices, in contrast to the known model of Danilevsky Yu.G., consisting of a single matrix of a complex structure. Secondly, the matrix elements will be the intensity characterizing the parameters of the workflow. This can significantly reduce the number of intermediate operations.

Having built this model, we can accurately determine the intensity of information interaction between all departments, the amount of information transmitted, and reasonably calculate the characteristics of data transmission channels. Knowing the structure of the flows across the enterprise and their ratio within each of the units, you can design the architecture of the information network, the structure of the data warehouse.

This research is funded by RFBR, grant 17-07-00096.

- [1] *Sviridov A.* Method of constructing a mathematical model of enterprise information flows // *Izvestiya UFU. Tekhnicheskiye nauki*, Taganrog: Publisher UFU, 2004. — p. 152–156.

Содержание

Теория и методы машинного обучения	10
<i>Бахтеев О. Ю.</i>	
Градиентные методы оптимизации гиперпараметров моделей глубокого обучения	10
<i>Бериков В. Б.</i>	
Построение оптимального коллективного решения в кластерном анализе на основе усредненной коассоциативной матрицы и индексов качества группировки	12
<i>Бочкарёв А. М., Софронов И. Л., Стрижов В. В.</i>	
Прогноз проницаемости горной породы с помощью символической регрессии	14
<i>Визильтер Ю. В., Горбачевич В. С.</i>	
Структурно-функциональный анализ и синтез глубоких конволюционных нейронных сетей	16
<i>Генрихов И. Е., Дюкова Е. В., Журавлёв В. И.</i>	
О полных решающих деревьях в задаче восстановления регрессии	18
<i>Гончаров А. В., Стрижов В. В.</i>	
Построение интерпретируемых моделей глубокого обучения в задаче социального ранжирования	20
<i>Двоенко С. Д., Пшеничный Д. О.</i>	
Исследование обусловленности матриц парных сравнений при их метрической коррекции	22
<i>Дюкова Е. В., Масляков Г. О., Прокофьев П. А.</i>	
О дуализации над производением цепей	24
<i>Исаченко Р. В., Жариков И. Н., Бочкарёв А. М.</i>	
Локальные модели для классификации объектов сложной структуры	26
<i>Красоткина О. В., Моттль В. В., Марков М.</i>	
Методы оценивания гиперпараметров в задачах динамического анализа стиля инвестиций	28

<i>Маленичев А. А., Моттль В. В., Красоткина О. В.</i> Быстрые последовательные методы обучения обобщенных линейных моделей зависимостей	30
<i>Моттль В. В., Середин О. С.</i> Научное наследие Эммануила Марковича Бравермана в методологии машинного обучения	32
<i>Неделько В. М.</i> Оценивание значимости переменных для ранговой регрессии	34
<i>Нейчев Р. Г., Стрижов В. В.</i> Построение модели оптимальной сложности в виде смеси экспертов	36
<i>Немирко А. П.</i> Сокращение размерности признакового пространства на основе критерия минимального расстояния между выпуклыми оболочками классов	38
<i>Пушняков А. С.</i> Об улучшенной оценке меры кластерной структуры в компактном метрическом пространстве	40
<i>Шибзухов З. М.</i> О принципе минимизации эмпирического риска на основе усредняющих агрегирующих функций	42
<i>Янковская А. Е., Ямианов А. В.</i> Интеллектуальная система поддержки принятия решений с когнитивной компонентой, основанная на отказоустойчивых диагностических тестах	44
Дискретная оптимизация и сложность вычислений	46
<i>Агеев А. А., Кельманов А. В., Пяткин А. В., Хамидуллин С. А., Шенмайер В. В.</i> Приближенный полиномиальный алгоритм для задачи очистки и редактирования данных	46
<i>Германчук М. С., Козлова М. Г., Лукьяненко В. А.</i> Задачи дискретной оптимизации типа многих коммивояжеров	48

<i>Гимади Э. Х.</i>	
О реализации асимптотически точного подхода к решению некоторых трудных комбинаторных задач	50
<i>Зухба А. В.</i>	
Алгоритм монотонизации выборки с одновременным отбором объектов и признаков	52
<i>Кельманов А. В.</i>	
О некоторых задачах кластеризации	54
<i>Лемтюжникова Д. В.</i>	
Декомпозиция разреженных матриц в задачах целочисленного программирования	56
<i>Хачай М. Ю., Незнакина Е. Д.</i>	
Эффективная разрешимость обобщенной задачи коммивояжера в классах квази- и псевдопирамидальных маршрутов	58
<i>Хачай М. Ю., Панкратов В. С.</i>	
Гарантированная оценка качества кластеризации методом k -медиан на отрезке	60
Анализ сигналов и временных рядов	62
<i>Бабичев Д. О., Красоткина О. В., Моттль В. В.</i>	
Отбор подмножества активных биржевых индексов в динамической модели инвестиционного портфеля	62
<i>Кирилюк И. Л., Кузнецова А. В., Сенько О. В.</i>	
Оптимизация сложности моделей анализа временных рядов в экономике	64
<i>Мангалова Е. С.</i>	
Вероятностное прогнозирование мощности ветряных электростанций с использованием алгоритма К ближайших соседей	66
<i>Пугач И. А., Моттль В. В., Красоткина О. В.</i>	
Алгоритмы анализа динамики стиля инвестиций по большому числу биржевых индексов	68

<i>Рудаков К. В., Кузнецов М. К., Мотренко А. П., Громов А. Н., Попова Е. В., Стрижов В. В.</i> Выбор оптимальной модели прогнозирования объемов грузовых железнодорожных перевозок	70
<i>Старожилец В. М., Чехович Ю. В.</i> Об идентификации статистической модели транспортных потоков с использованием групп АТС	72
<i>Филипенков Н. В., Петрова М. А.</i> Поиск плавно меняющихся закономерностей на рынке недвижимости	74
Обработка изображений	76
<i>Архипова О. Е., Герасюк В. С.</i> Использование методов обработки космических изображений для оценки состояния рекреационной зоны Таганрогского залива	76
<i>Архипова О. Е., Сурков Ф. А.</i> Представление больших данных в задачах поддержки принятия решений по обеспечению устойчивого развития региона	78
<i>Грачева И. А., Копылов А. В.</i> Алгоритм обработки изображений на основе диагональной аппроксимации графа смежности элементов изображения	80
<i>Князь В. В., Выголов О. В., Федоренко В. В., Севрюков В. Д.</i> Глубокие свёрточные автоэнкодеры: стереотождественные для восстановления 3D-моделей слабо текстурированных объектов	82
<i>Мурашов Д. М., Мурашов Ф. Д.</i> Метод локализации областей интереса с текстурой специального вида на изображениях	84
<i>Нестеренко В. А.</i> Поиск фоновых областей в изображении	86
<i>Титов В. С., Мирошниченко С. Ю.</i> Построение цифровых моделей рельефа по растровым топографическим картам	88

<i>Ханыков И. Г., Харинов М. В.</i> Модель применения метода Уорда для кластеризации пикселей изображения	90
<i>Шуклин И. И., Мосин С. А., Мирошниченко С. Ю.</i> Влияние свойств данных дистанционного зондирования земли и параметров их трансформирования на время построения результирующего изображения местности	92
Анализ и распознавание изображений	94
<i>Ефимов Ю. С., Одиноких Г. А., Соломатин И. А.</i> Разработка метода защиты от высококачественных подделок для мобильной системы распознавания по радужной оболочке	94
<i>Ланге М. М., Ганебных С. Н.</i> Комплексирование данных для классификации объектов по ансамблю изображений	96
<i>Моисеенко А. С., Горбачевич В. С., Визильтер Ю. В.</i> Полноконволюционная сеть FaceDetectNet для обнаружения лиц	98
<i>Савченко А. В.</i> Распознавание изображений на основе метода максимально правдоподобных рассогласований	100
<i>Самсонов Н. А., Гнеушев А. Н.</i> Дескриптор в аккумуляторном пространстве Хафа градиентного поля изображения для детектирования пешеходов	102
<i>Федотов Н. Г., Сёмов А. А., Моисеев А. В.</i> Исследование быстродействия распознавания 3D изображения методами стохастической геометрии в зависимости от количества опорных точек на сфере	104
<i>Чернышов В. Г.</i> Метод биометрической идентификации по изображению внешней стороны ладони	106
<i>Чигринский В. В., Матвеев И. А.</i> Исследование движения структуры радужной оболочки глаза методом оптического потока	108

Морфология изображений	110
<i>Брянский С. А., Визильтер Ю. В.</i>	
Условные морфологические меры сложности и информативности изображений	110
<i>Жукова К. В., Рейер И. А., Инякин А. С.</i>	
Поиск изображений по визуальным характеристикам с использованием гранично-скелетной модели формы	112
<i>Кушнир О. А., Середин О. С.</i>	
Распараллеливание алгоритма сравнения бинарных изображений, представленных цепочками скелетных примитивов	114
<i>Лебедев М. А., Рубис А. Ю., Визильтер Ю. В., Выголов О. В., Желтов С. Ю.</i>	
Выделение отличий на изображениях с помощью референтных EMD-фильтров	116
<i>Ломов Н. А., Сидякин С. В.</i>	
Классификация бинарных фигур с использованием мешка контекстов ребер скелета	118
<i>Местецкий Л. М., Тирас Х. П.</i>	
Морфологическая оценка динамики регенерации планарий по фотоизображениям	120
<i>Мирошниченко С. Ю., Дремов Е. Н.</i>	
Метод выделения подмножества векторных представлений утолщенных горизонталей на растровых топографических картах	122
<i>Перевалов Д. С., Корнилов Ф. А.</i>	
Морфологический проектор в метрике L_0 и задача поиска структурных различий изображений	124
<i>Рубис А. Ю., Лебедев М. А., Визильтер Ю. В.</i>	
Компаративная фильтрация изображений с использованием монотонных морфологических операторов	126
Анализ текстов и информационный поиск	128
<i>Бахтеев О. Ю., Кузнецова М. В.</i>	
Детектирование переводных заимствований в текстах научных статей из журналов, входящих в РИНЦ	128

<i>Воронцов К. В.</i>	
Аддитивная регуляризация тематических моделей связанного текста	130
<i>Копылов А. В., Середин О. С., Найденов А. В., Зенин Д. Г.</i>	
Формирование базы данных для системы оценки эмоционального фона диалога с оператором центра обработки вызовов	132
<i>Липко Ю. Ю., Свиридов А. С.</i>	
Исследование и реализация рекомендательных веб-сервисов выбора образовательных ресурсов для реализации индивидуальных образовательных траекторий, обучающихся на основе коллаборативной фильтрации	134
<i>Михайлов Д. В., Козлов А. П., Емельянов Г. М.</i>	
Мера TF-IDF, сила связи слов и формирование единиц представления знаний в открытых тестах	136
<i>Мохов А. С., Толчеев В. О.</i>	
Использование структуры научных документов в задачах обработки и анализа текстовой информации	138
<i>Новиков Д. Д.</i>	
Метод семантических ядер для классификации публичных чатов мессенджера Telegram	140
<i>Рогозов Ю. И.</i>	
Парадигма смыслового проектирования онтологий	142
Анализ биомедицинских сигналов	144
<i>Анижеев Д. А., Пенкин Г. О., Стрижов В. В.</i>	
Классификация физической активности человека с помощью локальных аппроксимирующих моделей	144
<i>Буторина А. В., Павлова А. А.</i>	
Локализация источников мозговой активности по МЭГ данным в исследованиях семантического прайминга	146
<i>Кершнер И. А., Комольцев И. Г., Обухов К. Ю., Обухов Ю. В.</i>	
Автоматическое детектирование событий в долговременных записях ЭЭГ	148

<i>Мачило Л. А., Немирко А. П.</i> Анализ биомедицинских сигналов с нелинейными свойствами на основе энтропийных характеристик	150
<i>Морозов А. А., Сушкова О. С., Полупанов А. Ф.</i> Разработка методов и алгоритмов анализа видеоизображений лабораторных животных средствами объектно-ориентированного логического программирования	152
<i>Мотренко А. П.</i> Выбор признаков в задаче декодирования сигналов ECoG	154
<i>Оплачко Е. С., Рыкунов С. Д., Устинин М. Н.</i> Mathbrain.ru — облачный сервис для обработки данных магнитной энцефалографии	156
<i>Рыкунов С. Д., Сычев В. В., Молчанова Д. А., Устинин М. Н.</i> Нахождение парциальных спектров разделов головного мозга методом анализа в пространстве частота-паттерн	158
<i>Сарманова О. Э., Буриков С. А., Доленко С. А., Исаев И. В., Светлов В. А., Лаптинский К. А., Доленко Т. А.</i> Методы машинного обучения в решении задачи мониторинга выведения тераностических флуоресцентных наноконструктов из организма	160
<i>Сушкова О. С., Морозов А. А., Габова А. В.</i> Исследование специфичности признаков раннего паркинсонизма, полученных методом анализа всплескообразной электрической активности коры мозга	162
<i>Устинин М. Н., Рыкунов С. Д., Бойко А. И., Сычев В. В., Молчанова Д. А.</i> Реконструкция функциональной структуры сложных систем по многоканальным временным рядам	164
Биоинформатика	166
<i>Доленко Т. А., Буриков С. А., Вервальд Е. Н., Ефиторов А. О., Лаптинский К. А., Сарманова О. Э., Доленко С. А.</i> Повышение достоверности молекулярных ДНК-вычислений: решение обратной задачи спектроскопии комбинационного рассеяния с помощью искусственных нейронных сетей	166

<i>Панкратов А. Н.</i> Спектрально-аналитический метод распознавания неточных повторов в символьных последовательностях	168
<i>Тетуев Р. К., Пятков М. И., Панкратов А. Н.</i> Параллельный алгоритм глобального выравнивания протяжённых аминокислотных и нуклеотидных последовательностей	170
<i>Тихонов Д. А., Куликова Л. И., Ефимов А. В.</i> Исследование межспиральных углов в структурных мотивах, образованных двумя спиралями	172
Геоинформатика	174
<i>Горнов А. Ю., Берзинский Ю. А., Финкельштейн Е. А., Зароднюк Т. С., Ордынская А. П.</i> Оценка пластического ресурса каркаса критически важного объекта при воздействии максимального расчетного землетрясения	174
<i>Флоринский И. В.</i> Моделирование рельефа на поверхности сферы, эллипсоида вращения и трехосного эллипсоида	176
Прикладные системы	178
<i>Астафьев А. В., Орлов А. А.</i> Разработка алгоритма контроля перемещения изделий между зонами складирования на основе статистического анализа временных рядов данных с их идентификаторов для систем промышленной автоматизации	178
<i>Жарких А. А., Червяков И. О.</i> Перспективы использования кодов с коррекцией ошибок в стеганографических системах, использующих в качестве контейнеров аудио файлы	180
<i>Каржищенко А. Н., Мнухин В. Б.</i> Применение гауссовских вращений для защиты графической информации	182

<i>Руднев Д. О., Сычугов А. А.</i>	
Методы построения систем анализа состояния распределённых информационных систем на основе беспризнакового распознавания образов	184
<i>Токарев В. Л., Сычугов А. А.</i>	
Обнаружение вредоносного программного обеспечения с использованием иммунных детекторов	186
Программная инженерия	188
<i>Беликов А. Н.</i>	
Абстракция метода как средства создания понятий	188
<i>Беликова С. А.</i>	
Проектирование пользовательского интерфейса ИС на основе семантического подхода	190
<i>Грищенко А. С.</i>	
Разработка метода построения процедур вставки данных в структурно-независимые базы данных на основе деятельностного подхода	192
<i>Елькин Д. М.</i>	
Разработка новых методов организации дорожного движения на основе самоорганизации связей	194
<i>Кучеров С. А., Rogozov Ю. И.</i>	
NoSQL-подход к хранению данных в конфигурируемой информационной системе	196
<i>Лапшин В. С.</i>	
Разработка механизмов трансляции программных моделей, представленных в формате среды Matlab	198
<i>Свиридов А. С., Липко Ю. Ю.</i>	
Модель информационных потоков для проектирования информационной сети предприятия	200

Contents

Machine Learning	10
<i>Bakhteev O.</i>	
Gradient-based hyperparameter optimization for deep learning models	11
<i>Berikov V.</i>	
Construction of an Optimal Collective Decision in Cluster Analysis on the Basis of an Averaged Co-Association Matrix and Cluster Validity Indices	13
<i>Bochkarev A., Sofronov I., Strijov V.</i>	
Permeability prediction using symbolic regression	15
<i>Vizilter Yu., Gorbatshevich V.</i>	
Structural-Functional Analysis and Synthesis of Deep Convolutional Neural Networks	17
<i>Genrikhov I., Djukova E., Zhuravlyov V.</i>	
On full decision trees in regression restoration problem	19
<i>Goncharov A., Strijov V.</i>	
Deep interpreted models in social ranking task	21
<i>Dvoenko S., Pshenichny D.</i>	
Investigation of matrix conditionality under metric correction of pairwise comparisons	23
<i>Djukova E., Maslyakov G., Prokofyev P.</i>	
About dualization over product of chains	25
<i>Isachenko R., Zharikov I., Bochkarev A.</i>	
Local models for classification of complex structured objects	27
<i>Krasotkina O., Mottl V., Markov M.</i>	
Methods of hyperparameter estimation in time-varying regression models with application to dynamic style analysis of investment portfolios	29
<i>Makenichev A., Mottl V., Krasotkina O.</i>	
Algorithms for dynamic style analysis of investments from a large number of stock market indices	31

<i>Mottl V., Seredin O.</i>	
The scientific heritage of Emmanuel Braverman in the methodology of Machine Learning	33
<i>Nedel'ko V.</i>	
Estimation of feature importance for quantile regression . . .	35
<i>Neychev R., Strijov V.</i>	
Mixture of Experts approach to model of optimal complexity construction	37
<i>Nemirko A.</i>	
Reducing the dimension of feature space based on the crite- rion of minimum distance between the convex hulls	39
<i>Pushnyakov A.</i>	
On improved bound for measure of cluster structure in com- pact metric spaces	41
<i>Shibzukhov Z.</i>	
On the principle of empirical risk minimization based on averaging aggregation functions	43
<i>Yankovskaya A., Yamshanov A.</i>	
Intelligent system of decision-making support with cognitive component based on the fault-tolerant diagnostic tests	45
Discrete Optimization and Computational Complexity	46
<i>Ageev A., Kel'manov A., Pyatkin A., Khamidullin S., Shen- maier V.</i>	
Approximation Polynomial Algorithm for the Data Editing and Data Cleaning Problem	47
<i>Germanchuk M., Kozlova M., Lukianenko V.</i>	
The problems of discrete optimization such as Multiple Trav- eling Salesman Problems	49
<i>Gimadi E.</i>	
On Asymptotically Optimal Approach to Solving Some Hard Combinatorial Problems	51
<i>Zukhba A.</i>	
Monotonization algorithm with simultaneous feature and prototype selection	53

<i>Kel'manov A.</i>	
On some clustering problems	55
<i>Lemtyuzhnikova D.</i>	
Decomposition of sparse matrices in integer programming problems	57
<i>Khachay M., Neznakhina K.</i>	
Efficient solvability of the Quasi- and Pseudo-Pyramidal Generalized Traveling Salesman Problem	59
<i>Khachay M., Pankratov V.</i>	
Attainable best accuracy guarantee for k -medians clustering in $[0, 1]$	61
Signal and Time Series Analysis	62
<i>Babichev D., Krasotkina O., Mottl V.</i>	
Selection of the subset of active stock market indices in the dynamic model of an investment portfolio	63
<i>Kirilyuk I., Kuznetsova A., Senko O.</i>	
Optimization of models complexity for time series analysis in economics	65
<i>Mangalova E.</i>	
Probabilistic wind power forecasting using K-nearest neigh- bors algorithm	67
<i>Pugach I., Mottl V., Krasotkina O.</i>	
Algorithms for dynamic style analysis of investments from a large number of stock market indices	69
<i>Rudakov K., Kuznetsov M., Motrenko A., Gromov A., Popova E., Strijov V.</i>	
On selecting the optimal forecasting model for railway freight transportation demand	71
<i>Starozhilets V., Chehovich Yu.</i>	
About identification of a statistical model of traffic flows using vehicle groups	73
<i>Filipenkov N., Petrova M.</i>	
Mining Slightly Changing Patterns on the Housing Market .	75

Image Processing	76
<i>Arkhipova O., Gerasyuk V.</i>	
Space-imaging techniques to assess the status of a recreational zone of the Taganrog Bay	77
<i>Arkhipova O., Surkov F.</i>	
Representation of large data in problems of decision support for sustainable development of the region	79
<i>Gracheva I., Kopylov A.</i>	
An image processing algorithm based on the diagonal approximation of the adjacency graph of image elements	81
<i>Kniaz V., Vygolov O., Fedorenko V., Sevrykov V.</i>	
Deep convolutional autoencoders: stereo matching for 3D model reconstruction of low-textured objects	83
<i>Murashov D., Murashov F.</i>	
Method for localizing image regions of interest with a specific texture	85
<i>Nesterenko V.</i>	
The search for background areas in an image	87
<i>Titov V., Miroshnichenko S.</i>	
Digital elevation models creation from raster topographic maps	89
<i>Khanykov I., Kharinov M.</i>	
A model for implementation of Ward's method of image pixel clustering	91
<i>Shuklin I., Mosin S., Miroshnichenko S.</i>	
Earth remote sensing data features and transformation parameters effect on terrain image creation process duration	93
Image Analysis and Recognition	94
<i>Efimov Yu., Odinokih G., Solomatin I.</i>	
High-quality presentation attack detection in a mobile iris recognition system	95
<i>Lange M., Ganebnykh S.</i>	
Data Fusion Schemes for Object Classification Using Ensemble of Images	97

<i>Moiseenko A., Vizilter Yu., Gorbatsевич V.</i>	
FaceDetectNet Fully-convolutional Network for Face Detection	99
<i>Savchenko A.</i>	
Image recognition based on the method of maximum likelihood dissimilarities	101
<i>Samsonov N., Gbeushev A.</i>	
Textural descriptor in the Hough accumulator space of the gradient field for detecting pedestrians	103
<i>Fedotov N., Syemov A., Moiseev A.</i>	
Performance investigation of 3D image recognition by stochastic geometry methods in dependent on the number of reference points on the sphere	105
<i>Chernyshov V.</i>	
Biometric identification using hand dorsal surface	107
<i>Chigrinskiy V., Matveev I.</i>	
Iris structure motion analysis via optical flow method	109
Morphological Image Processing	110
<i>Brianskiy S., Vizilter Yu.</i>	
Morphology measures of complexity and information for binary images	111
<i>Zhukova K., Reyer I., Inyakin A.</i>	
Region-based image retrieval with use of boundary-skeleton shape model	113
<i>Kushnir O., Seredin O.</i>	
The Parallel Implementations of the Shape Comparison Algorithm Based on Skeleton Primitive Chains Alignment	115
<i>Lebedev M., Rubis A., Vizilter Yu., Vygolov O., Zheltov S.</i>	
Image differences detection based on reference EMD-filters	117
<i>Lomov N., Sidyakin S.</i>	
Shape classification using bag of edge contexts	119
<i>Mestetskiy L., Tiras K.</i>	
Morphological evaluation of the dynamics of planarian regeneration through photo-images	121

<i>Miroshnichenko S., Dremov E.</i>	
Method selection subset of the vector representation thickening horizontals on raster topographic maps	123
<i>Perevalov D., Kornilov F.</i>	
Morphological projector in the L_0 metric and the problem of the finding structural differences between images	125
<i>Rubis A., Levedev M., Vizilter Yu.</i>	
Comparative image filtering using monotonical morphology operators	127
Text Analysis and Information Retrieval	128
<i>Bakhteev O., Kuznetsova R.</i>	
Cross-lingual plagiarism detection in the collection of scientific articles from journals included in RSI	129
<i>Vorontsov K.</i>	
Additive regularization for topic modeling of coherent text .	131
<i>Kopylov A., Seredin O., Naidyonov A., Zenin D.</i>	
The creation of a corpus of emotional data for the system of emotion-related states assessment of a dialogue with the call center operator	133
<i>Lipko Yu., Sviridov A.</i>	
Research and implementation of recommended web services for the selection of educational resources for the implementation of individual educational trajectories of students on the basis of collaborative filtering	135
<i>Mikhaylov D., Kozlov A., Emelyanov G.</i>	
TF-IDF, coupling strength of words and formation of units for knowledge representation in open tests	137
<i>Mokhov A., Tolcheev V.</i>	
Usage of the structure of scientific documents in tasks of processing and analysis of text information	139
<i>Novikov D.</i>	
Semantic kernels method for Telegram messenger public chats classification	141

<i>Rogozov Yu.</i>	
The paradigm of the semantic design of ontologies	143
Biomedical Signal Analysis	144
<i>Anikeyev D., Penkin P., Strijov V.</i>	
Local approximation for human physical activity classification	145
<i>Butorina A., Pavlova A.</i>	
MEG source localization in studies of semantic priming . . .	147
<i>Kershner I., Komoltsev I., Obukhov K., Obukhov Yu.</i>	
Approach for EEG of Post Traumatic Sleep Spindles and Epilepsy Seizures Detection and Classification in Rats . . .	149
<i>Manilo L., Nemirko A.</i>	
Analysis of biomedical signals with nonlinear properties based on entropy characteristics	151
<i>Morozov A., Sushkova O., Polupanov A.</i>	
Development of methods and algorithms based on object- oriented logic programming for video monitoring of labora- tory animals	153
<i>Motrenko A.</i>	
Multi-way Feature Selection for ECoG-based Brain-Computer Interface	155
<i>Oplachko E., Rykunov S., Ustinin M.</i>	
Mathbrain.ru — a cloud service for magnetic encephalogra- phy data processing	157
<i>Rykunov S., Sychev V., Molchanova D., Ustinin M.</i>	
Estimation of the brain compartments partial spectra by frequency-pattern analysis	159
<i>Sarmanova O., Burikov S., Dolenko S., Isaev I., Svetlov V., Laptinskiy K., Dolenko T.</i>	
Machine learning methods for the purpose of monitoring of the excretion of theranostic fluorescent nanocomposites out of the organism	161
<i>Sushkova O., Morozov A., Gabova A.</i>	
An investigation of specificity of features of early stages of Parkinson's disease obtained using the method of cerebral cortex electrical activity analysis based on wave trains . . .	163

<i>Ustinin M., Rykunov S., Boyko A., Sychev V., Molchanova D.</i> Reconstruction of the functional structure of complex systems from the multichannel time series	165
Bioinformatics	166
<i>Dolenko T., Burikov S., Vervalde E., Efitorov A., Laptinskiy K., Sarmanova O., Dolenko S.</i> Improvement of reliability of molecular DNA computing: solution of inverse problem of Raman spectroscopy using artificial neural networks	167
<i>Pankratov A.</i> Spectral Analytical Method of Recognition of Inexact Repeats in Character Sequences	169
<i>Tetuev R., Pyatkov M., Pankratov A.</i> Parallel algorithm for global alignment of long aminoacid and nucleotide sequences	171
<i>Tikhonov D., Kulikova L., Efimov A.</i> The study of interhelical angles in the structural motifs formed by two helices	173
Geoinformatics	174
<i>Gornov A., Berzhinsky Yu., Finkelstein E., Zarodnyuk T., Ordynskaya A.</i> The estimation for the plastic resource of the framework of a critically important object under the action of the maximum design earthquake	175
<i>Florinsky I.</i> Terrain modelling on a surface of a sphere, ellipsoid of revolution, and triaxial ellipsoid	177
Applied Systems	178
<i>Astafiev A., Orlov A.</i> Development of an algorithm for controlling the movement of products between storage areas based on the statistical analysis of time series data from their identifiers for industrial automation systems	179

<i>Zharkikh A., Chervyakov I.</i>	
Prospects for the use of error correction codes in steganographic systems using audio files as containers	181
<i>Karkishchenko A., Mnukhin V.</i>	
Gaussian rotations for graphic information protection	183
<i>Rudnev D., Sychugov A.</i>	
Methods for constructing systems for analyzing the state of distributed information systems based on unrecognized pattern recognition	185
<i>Tokarev V., Sychugov A.</i>	
Detection of malicious software using immune detectors	187
Software Engineering	188
<i>Belikov A.</i>	
The abstraction of a method as means of creating concepts	189
<i>Belikova S.</i>	
Information systems user interface design based on semantic approach	191
<i>Grishchenko A.</i>	
Method of creating procedures for inserting data in structure-independent databases based on activity driven approach	193
<i>Elkin D.</i>	
Development of new methods for organizing traffic on the basis of self-organization of links	195
<i>Kucherov S., Rogozov Yu.</i>	
NoSQL Approach to Data Storing in Configurable Information Systems	197
<i>Lapshin V.</i>	
Development of mechanisms for translating software models presented in the matlab environment format	199
<i>Sviridov A., Lipko Yu.</i>	
Model of information flows for the design of enterprise information network	201

Авторский указатель**А**

Агеев А. А. 46
Аникеев Д. А. 144
Аршипова О. Е. 76, 78
Астафьев А. В. 178

Б

Бабичев Д. О. 62
Бахтеев О. Ю. 10, 128
Беликов А. Н. 188
Беликова С. А. 190
Бержинский Ю. А. 174
Бериков В. Б. 12
Бойко А. И. 164
Бочкарёв А. М. 14, 26
Брянский С. А. 110
Буриков С. А. 160, 166
Буторина А. В. 146

В

Вервальд Е. Н. 166
Визильтер Ю. В. 16, 98, 110,
116, 126
Воронцов К. В. 130
Выголов О. В. 82, 116

Г

Габова А. В. 162
Ганебных С. Н. 96
Генрихов И. Е. 18
Герасюк В. С. 76
Германчук М. С. 48
Гимади Э. Х. 50
Гнеушев А. Н. 102
Гончаров А. В. 20
Горбацевич В. С. 16, 98

Горнов А. Ю. 174
Грачева И. А. 80
Грищенко А. С. 192
Громов А. Н. 70

Д

Двоенко С. Д. 22
Доленко С. А. 160, 166
Доленко Т. А. 160, 166
Дремов Е. Н. 122
Дюкова Е. В. 18, 24

Е

Елькин Д. М. 194
Емельянов Г. М. 136
Ефимов А. В. 172
Ефимов Ю. С. 94
Ефиторов А. О. 166

Ж

Жариков И. Н. 26
Жарких А. А. 180
Желтов С. Ю. 116
Жукова К. В. 112
Журавлёв В. И. 18

З

Зароднюк Т. С. 174
Зенин Д. Г. 132
Зухба А. В. 52

И

Инякин А. С. 112
Исаев И. В. 160
Исаченко Р. В. 26

К

Каркищенко А. Н. 182
 Кельманов А. В. 46, 54
 Кершнер И. А. 148
 Кириллук И. Л. 64
 Князь В. В. 82
 Козлов А. П. 136
 Козлова М. Г. 48
 Комольцев И. Г. 148
 Копылов А. В. 80, 132
 Корнилов Ф. А. 124
 Красоткина О. В. 28, 30, 62,
 68
 Кузнецов М. К. 70
 Кузнецова А. В. 64
 Кузнецова М. В. 128
 Куликова Л. И. 172
 Кучеров С. А. 196
 Кушнир О. А. 114

Л

Ланге М. М. 96
 Лаптинский К. А. .. 160, 166
 Лапшин В. С. 198
 Лебедев М. А. 116, 126
 Лемтюжникова Д. В. 56
 Липко Ю. Ю. 134, 200
 Ломов Н. А. 118
 Лукьяненко В. А. 48

М

Маленичев А. А. 30
 Мангалова Е. С. 66
 Манило Л. А. 150
 Марков М. 28
 Масляков Г. О. 24
 Матвеев И. А. 108
 Местецкий Л. М. 120

Мирошниченко С. Ю. ... 88,
 92, 122
 Михайлов Д. В. 136
 Мнухин В. Б. 182
 Моисеев А. В. 104
 Моисеенко А. С. 98
 Молчанова Д. А. ... 158, 164
 Морозов А. А. 152, 162
 Мосин С. А. 92
 Мотренко А. П. 70, 154
 Моттль В. В. . 28, 30, 32, 62,
 68
 Мохов А. С. 138
 Мурашов Д. М. 84
 Мурашов Ф. Д. 84

Н

Найденов А. В. 132
 Неделько В. М. 34
 Незнахина Е. Д. 58
 Нейчев Р. Г. 36
 Немирко А. П. 38, 150
 Нестеренко В. А. 86
 Новиков Д. Д. 140

О

Обухов К. Ю. 148
 Обухов Ю. В. 148
 Одиноких Г. А. 94
 Оплачко Е. С. 156
 Ордынская А. П. 174
 Орлов А. А. 178

П

Павлова А. А. 146
 Панкратов А. Н. ... 168, 170
 Панкратов В. С. 60
 Пенкин Г. О. 144

Перевалов Д. С. 124
Петрова М. А. 74
Полупанов А. Ф. 152
Попова Е. В. 70
Прокофьев П. А. 24
Пугач И. А. 68
Пушняков А. С. 40
Пшеничный Д. О. 22
Пяткин А. В. 46
Пятков М. И. 170

Р

Рейер И. А. 112
Рогозов Ю. И. 142, 196
Рубис А. Ю. 116, 126
Рудаков К. В. 70
Руднев Д. О. 184
Рыкунов С. Д. 156, 158, 164

С

Сёмов А. А. 104
Савченко А. В. 100
Самсонов Н. А. 102
Сарманова О. Э. 160, 166
Светлов В. А. 160
Свиридов А. С. 134, 200
Севрюков В. Д. 82
Сенько О. В. 64
Середин О. С. 32, 114, 132
Сидякин С. В. 118
Соломатин И. А. 94
Софронов И. Л. 14
Старожилец В. М. 72
Стрижов В. В. 14, 20, 36, 70,
144
Сурков Ф. А. 78
Сушкова О. С. 152, 162
Сычев В. В. 158, 164

Сычугтов А. А. 184, 186

Т

Тегуев Р. К. 170
Тирас Х. П. 120
Титов В. С. 88
Тихонов Д. А. 172
Токарев В. Л. 186
Толчеев В. О. 138

У

Устинин М. Н. 156, 158, 164

Ф

Федоренко В. В. 82
Федотов Н. Г. 104
Филипенков Н. В. 74
Финкельштейн Е. А. 174
Флоринский И. В. 176

Х

Хамидуллин С. А. 46
Ханьков И. Г. 90
Харинов М. В. 90
Хачай М. Ю. 58, 60

Ч

Червяков И. О. 180
Чернышов В. Г. 106
Чехович Ю. В. 72
Чигринский В. В. 108

Ш

Шенмайер В. В. 46
Шибзухов З. М. 42
Шуклин И. И. 92

Я

Ямшанов А. В. 44
Янковская А. Е. 44

Author index

A

Ageev A. 47
 Anikeyev D. 145
 Arkhipova O. 77, 79
 Astafiev A. 179

B

Babichev D. 63
 Bakhteev O. 11, 129
 Belikova S. 191
 Belikov A. 189
 Berikov V. 13
 Berzhinsky Yu. 175
 Bochkarev A. 15, 27
 Boyko A. 165
 Brianskiy S. 111
 Burikov S. 161, 167
 Butorina A. 147

C

Chehovich Yu. 73
 Chernyshov V. 107
 Chervyakov I. 181
 Chigrinskiy V. 109

D

Djukova E. 19, 25
 Dolenko S. 161, 167
 Dolenko T. 161, 167
 Dremov E. 123
 Dvoenko S. 23

E

Efimov A. 173
 Efimov Yu. 95
 Efitorov A. 167

Elkin D. 195
 Emelyanov G. 137

F

Fedorenko V. 83
 Fedotov N. 105
 Filipenkov N. 75
 Finkelstein E. 175
 Florinsky I. 177

G

Gabova A. 163
 Ganebnykh S. 97
 Gbeushev A. 103
 Genrikhov I. 19
 Gerasyuk V. 77
 Germanchuk M. 49
 Gimadi E. 51
 Goncharov A. 21
 Gorbatshevich V. 17, 99
 Gornov A. 175
 Gracheva I. 81
 Grishchenko A. 193
 Gromov A. 71

I

Inyakin A. 113
 Isachenko R. 27
 Isaev I. 161

K

Karkishchenko A. 183
 Kel'manov A. 47, 55
 Kershner I. 149
 Khachay M. 59, 61
 Khamidullin S. 47

Khanykov I. 91
Kharinov M. 91
Kirilyuk I. 65
Kniaz V. 83
Komoltsev I. 149
Kopylov A. 81, 133
Kornilov F. 125
Kozlova M. 49
Kozlov A. 137
Krasotkina O. . 29, 31, 63, 69
Kucherov S. 197
Kulikova L. 173
Kushnir O. 115
Kuznetsova A. 65
Kuznetsova R. 129
Kuznetsov M. 71

L

Lange M. 97
Lapshin V. 199
Laptinskiy K. 161, 167
Lebedev M. 117
Lemtyuzhnikova D. 57
Levedev M. 127
Lipko Yu. 135, 201
Lomov N. 119
Lukianenko V. 49

M

Makenichev A. 31
Mangalova E. 67
Manilo L. 151
Markov M. 29
Maslyakov G. 25
Matveev I. 109
Mestetskiy L. 121
Mikhaylov D. 137
Miroshnichenko S. 89, 93, 123

Mnukhin V. 183
Moiseenko A. 99
Moiseev A. 105
Mokhov A. 139
Molchanova D. 159, 165
Morozov A. 153, 163
Mosin S. 93
Motrenko A. 71, 155
Mottl V. ... 29, 31, 33, 63, 69
Murashov D. 85
Murashov F. 85

N

Naidyonov A. 133
Nedel'ko V. 35
Nemirko A. 39, 151
Nesterenko V. 87
Neychev R. 37
Neznakhina K. 59
Novikov D. 141

O

Obukhov K. 149
Obukhov Yu. 149
Odinokinh G. 95
Oplachko E. 157
Ordynskaya A. 175
Orlov A. 179

P

Pankratov A. 169, 171
Pankratov V. 61
Pavlova A. 147
Penkin P. 145
Perevalov D. 125
Petrova M. 75
Polupanov A. 153
Popova E. 71

Prokofyev P.25
 Pshenichny D.23
 Pugach I.69
 Pushnyakov A.41
 Pyatkin A.47
 Pyatkov M.171

R

Reyer I.113
 Rogozov Yu.143, 197
 Rubis A.117, 127
 Rudakov K.71
 Rudnev D.185
 Rykunov S.157, 159, 165

S

Samsonov N.103
 Sarmanova O.161, 167
 Savchenko A.101
 Senko O.65
 Seredin O.33, 115, 133
 Sevrykov V.83
 Shenmaier V.47
 Shibzukhov Z.43
 Shuklin I.93
 Sidyakin S.119
 Sofronov I.15
 Solomatin I.95
 Starozhilets V.73
 Strijov V. .15, 21, 37, 71, 145
 Surkov F.79
 Sushkova O.153, 163
 Svetlov V.161
 Sviridov A.135, 201

Sychev V.159, 165
 Sychugov A.185, 187
 Syemov A.105

T

Tetuev R.171
 Tikhonov D.173
 Tiras K.121
 Titov V.89
 Tokarev V.187
 Tolcheev V.139

U

Ustinin M.157, 159, 165

V

Vervalde E.167
 Vizilter Yu. .17, 99, 111, 117,
 127
 Vorontsov K.131
 Vygolov O.83, 117

Y

Yamshanov A.45
 Yankovskaya A.45

Z

Zarodnyuk T.175
 Zenin D.133
 Zharikov I.27
 Zharkikh A.181
 Zheltov S.117
 Zhukova K.113
 Zhuravlyov V.19
 Zukhba A.53

Научное издание

МАТЕМАТИЧЕСКИЕ МЕТОДЫ
РАСПОЗНАВАНИЯ ОБРАЗОВ

Тезисы докладов
18-й Всероссийской конференции
с международным участием

Напечатано с готового оригинал-макета

Сдано в набор 22.07.17. Подписано в печать 02.10.17.
Формат 60×90/16. Бумага офсетная. Печать цифровая.
Усл.-печ. л. 14,25. Уч.-изд. л. 11,0. Тираж 250 экз.

Заказ № 997.

Издательство «ТОРУС ПРЕСС»
121614, г. Москва, ул. Крылатская 29-1-43
e-mail: torus@torus-press.ru
<http://www.torus-press.ru>



Отпечатано в НИПКЦ «Восход-А» с готовых файлов
Москва 109052, ул. Смирновская, д. 25, стр. 3

ISBN 978-5-94588-225-6



