

РОССИЙСКАЯ АКАДЕМИЯ НАУК
ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
«ИНФОРМАТИКА И УПРАВЛЕНИЕ» РАН
НАУЧНО-КООРДИНАЦИОННЫЙ ЦЕНТР
«ЦИФРОВЫЕ МЕТОДЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ»

ПРИ ПОДДЕРЖКЕ

РОССИЙСКОГО ФОНДА ФУНДАМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ,
КОМПАНИЙ ФОРЕКСИС И ЦСПИР

Интеллектуализация обработки информации ИОИ-2018

Тезисы докладов
12-й Международной конференции
(Москва, Россия – Гаэта, Италия)

ТОРУС МОСКВА
ПРЕСС  2018

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

ББК 22.1:32.973.26-018.2

И 73

Интеллектуализация обработки информации: Тезисы докладов 12-й Международной конференции (Москва, Россия–Гаэта, Италия, 2018). — М.: ТОРУС ПРЕСС, 2018. — 218 с.
ISBN 978-5-94588-236-2

В сборнике представлены тезисы докладов 12-й Международной конференции «Интеллектуализация обработки информации», проводимой Российской академией наук, Федеральным исследовательским центром «Информатика и управление» РАН, Научно-координационный центром «Цифровые методы интеллектуального анализа данных».

Конференция проводится с 1989 г., начиная с 2000 г. — регулярно один раз в два года, и является представительным научным форумом в области интеллектуального анализа данных, машинного обучения, распознавания образов, анализа изображений, обработки сигналов, дискретного анализа.

Организационный комитет ИОИ-2018 выражает особую благодарность компаниям Форексис и ЦСПиР, оказавшим неоценимую помощь при подготовке и проведении конференции. Конференция поддержана грантом РФФИ 18-07-20075.

Сайт конференции <http://mmro.ru>.

ББК 22.1:32.973.26-018.2

ISBN 978-5-94588-236-2

© Авторы докладов, 2018

© ФИЦ ИУ РАН, 2018

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

Intelligent Data Processing: Theory and Applications: Book of abstracts of the 12th International Conference (Moscow, Russia–Gaeta, Italy, 2018). — Moscow: TORUS PRESS, 2018. — 218 p. ISBN 978-5-94588-236-2

The volume contains the abstracts of the 12th International Conference «Intelligent Data Processing: Theory and Applications.» The conference is organized by the Russian Academy of Sciences, the Federal Research Center «Informatics and Control» of the Russian Academy of Sciences and the Scientific and Coordination Center «Digital Methods of Data Mining». The conference has been held biennially since 1989. 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 IDP-2018 is grateful to Forecsys Co. and CFRS Co. for providing assistance in the conference preparation and execution. The conference is funded by RFBR, grant 18-07-20075.

The conference website <http://mmro.ru/en/>.

BBK 22.1:32.973.26-018.2

ISBN 978-5-94588-236-2

© Authors of the abstracts, 2018

© FRC CSC RAS, 2018

Оргкомитет

Председатель: Журавлев Юрий Иванович, *академик РАН, ФИЦ ИУ РАН*

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

Громов Андрей Николаевич
Инякин Андрей Сергеевич, *к.ф.-м.н.*
Исаченко Роман Владимирович
Кальян Виктор Петрович
Лемтюжникова Дарья Владимировна, *к.ф.-м.н.*

Программный комитет

Председатель: Рудаков Константин Владимирович, *академик РАН, ФИЦ ИУ РАН*

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

Визильтер Юрий Валентинович, *д.ф.-м.н.*
Воронцов Константин Вячеславович, *д.ф.-м.н.*
Гимади Эдуард Хайрутдинович, *д.ф.-м.н.*
Горнов Александр Юрьевич, *д.т.н.*
Громова Ольга Алексеевна, *д.м.н.*
Двоенко Сергей Данилович, *д.ф.-м.н.*
Донской Владимир Иосифович, *д.ф.-м.н.*
Кельманов Александр Васильевич, *д.ф.-м.н.*
Местецкий Леонид Моисеевич, *д.т.н.*
Мотгль Вадим Вячеславович, *д.т.н.*
Осипов Геннадий Семенович, *д.ф.-м.н.*
Пытьев Юрий Петрович, *д.ф.-м.н.*
Рейер Иван Александрович, *к.т.н.*
Рязанов Владимир Васильевич, *д.ф.-м.н.*
Сойфер Виктор Александрович, *академик РАН*
Устинин Михаил Николаевич, *д.ф.-м.н.*
Хачай Михаил Юрьевич, *д.ф.-м.н.*
Чуличков Алексей Иванович, *д.ф.-м.н.*
Шананин Александр Алексеевич, *чл.-корр. РАН*

Organizing Committee

Chair: Yury Zhuravlev, *acad. of RAS*,
FRCCSC

Secretary: Yury Chekhovich, *C.Sc.*

Andrey Gromov
Roman Isachenko
Andrey Inyakin, *C.Sc.*
Viktor Kalian
Darya Lemtyuzhnikova, *C.Sc.*

Program Committee

Chair: Konstantin Rudakov, *acad. of RAS*,
FRCCSC

Secretary: Vadim Strijov, *D.Sc.*

Alexey Chulichkov, *D.Sc.*
Vladimir Donskoy, *D.Sc.*
Sergey Dvoenko, *D.Sc.*
Edward Gimadi, *D.Sc.*
Alexander Gornov, *D.Sc.*
Olga Gromova, *D.Sc.*
Alexander Kel'manov, *D.Sc.*
Michael Khachay, *D.Sc.*
Leonid Mestetskiy, *D.Sc.*
Vadim Mottl, *D.Sc.*
Gennady Osipov, *D.Sc.*
Yury Pytiev, *D.Sc.*
Ivan Reyer, *C.Sc.*
Vladimir Ryazanov, *D.Sc.*
Alexander Shananin, *corr. member of RAS*
Viktor Soyfer, *acad. of RAS*
Mikhail Ustinin, *D.Sc.*
Yury Vizilter, *D.Sc.*
Konstantin Vorontsov, *D.Sc.*

Рецензенты

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

Reviewers

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

Краткое оглавление

| | |
|---|-----|
| Теория и методы машинного обучения | 10 |
| Дискретная оптимизация и сложность вычислений | 42 |
| Анализ сигналов и временных рядов | 82 |
| Обработка изображений | 106 |
| Анализ и распознавание изображений | 126 |
| Морфология изображений | 138 |
| Анализ текстов и информационный поиск | 144 |
| Анализ биомедицинских сигналов | 156 |
| Биоинформатика | 166 |
| Геоинформатика | 170 |
| Прикладные системы | 182 |
| Программная инженерия | 186 |
| Содержание | 192 |

Brief contents

| | |
|--|-----|
| Machine Learning | 10 |
| Discrete Optimization and Computational Complexity | 42 |
| Signal and Time Series Analysis | 82 |
| Image Processing | 106 |
| Image Analysis and Recognition | 126 |
| Morphological Image Processing | 138 |
| Text Analysis and Information Retrieval | 144 |
| Biomedical Signal Analysis | 156 |
| Bioinformatics | 166 |
| Geoinformatics | 170 |
| Applied Systems | 182 |
| Software Engineering | 186 |
| Contents | 192 |

Модель и вычислительно эффективный метод ансамблевой кластеризации

Бериков Владимир Борисович^{1,2} berikov@math.nsc.ru
*Новиков Иван Александрович*² novikov.ivan.2015@gmail.com

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

²Новосибирск, Новосибирский государственный университет

DOI: 10.30826/IDP201801

В работе исследуется метод построения кластерного ансамбля на основе подхода, использующего усредненную с весами матрицу коассоциации. Элементы матрицы соответствуют парам объектов выборки и равны частотам отнесения пар в один и тот же кластер по всем вариантам разбиения. Формулируется вероятностная модель, с помощью которой находятся зависимости между характеристиками ансамбля и оценкой качества решения. Предложен способ нахождения оптимальных весов ансамбля, соответствующих вариантам разбиения. Кроме того, разработан вычислительно эффективный алгоритм построения ансамблевого решения (с близкой к линейной трудоемкостью), основанный на малоранговом представлении матрицы коассоциации в комбинации с алгоритмом спектрального кластерного анализа. Малоранговое представление матрицы позволяет ускорить вычисления за счет применения эффективных операций матричной алгебры, а также уменьшить объем требуемой памяти. Для поиска собственных чисел и собственных векторов Лапласиана, сформированного на основе матрицы коассоциации, применен алгоритм степенной итерации. Разработана также модификация алгоритма поиска связных компонент графа, с учетом специфики соответствующей ему матрицы коассоциации. С помощью численных экспериментов на тестовых и реальных задачах проведено исследование предложенного метода, показавшее его преимущество в точности и скорости вычислений, в сравнении с рядом аналогичных методов.

Работа поддержана грантом РФФИ № 18-07-600.

- [1] *Berikov V* A Probabilistic Model of Fuzzy Clustering Ensemble // Pattern Recognition and Image Analysis, 2018. Vol. 28(1). — P. 1–10.

Model and computationally efficient method of ensemble clustering

Berikov Vladimir^{1,2}

berikov@math.nsc.ru

*Novikov Ivan*²

novikov.ivan.2015@gmail.com

¹Novosibirsk, Sobolev Institute of mathematics SB RAS

²Novosibirsk, Novosibirsk State University

The method of constructing a cluster ensemble on the basis of the approach using the weighted averaged co-association matrix is investigated. The matrix elements correspond to the pairs of sample objects and are equal to the frequencies of assigning pairs to one and the same cluster for all variants of the partition. A probabilistic model is formulated, with the help of which there are obtained relationships between the characteristics of the ensemble and the estimates of the quality of the ensemble solution. The method of finding the optimal weights of the ensemble corresponding to the variants of the partition is proposed. In addition, a computationally efficient algorithm for constructing an ensemble solution (with near linear time complexity) was developed, based on the low-rank representation of the averaged co-association matrix in combination with Spectral Cluster Analysis. The low-rank representation of the matrix allows accelerating calculations by applying efficient matrix algebra operations, as well as reducing the amount of required memory. To find eigenvalues and eigenvectors of Laplacian defined on the basis of the averaged co-association matrix, Power Iteration algorithm was used. A modification of the algorithm for searching connected graph components has also been developed, taking into account the specifics of the corresponding averaged co-association matrix. With the help of numerical experiments on artificial and real problems, a study of the proposed method has been carried out, which showed its advantage in accuracy and speed of computation, in comparison with a number of similar methods.

This research is funded by RFBR, grant No 18-07-00600.

- [1] *Berikov V* A Probabilistic Model of Fuzzy Clustering Ensemble // Pattern Recognition and Image Analysis, 2018. Vol. 28(1). — P. 1–10.

Задача дуализации над произведением цепей: асимптотика типичного числа решений

Дюкова Елена Всеволодовна^{1*}

edjukova@mail.ru

*Масляков Глеб Олегович*²

gleb-mas@mail.ru

*Прокофьев Пётр Александрович*³

p_prok@mail.ru

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

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

³Москва, Институт машиноведения им. А. А. Благонравова РАН

DOI: 10.30826/IDP201802

Рассматривается одна из центральных труднорешаемых задач логического анализа данных — дуализация над произведением частичных порядков, которая формулируется следующим образом.

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 элементом $x = (x_1, \dots, x_n) \in P$, если y_i следует за x_i при $i = 1, 2, \dots, n$. Для обозначения того, что $y \in P$ следует за $x \in P$ и $y \neq x$ далее используется запись $x \prec y$. Пусть $R \subseteq P$, $R^+ = R \cup \{x \in P \mid \exists a \in P, a \prec x\}$. Задача построения двойственного к R множества $I(R)$, состоящего из элементов $a \in P \setminus R^+$ таких, что для любого $x \in P \setminus R^+$, $a \neq x$, отношение $a \prec x$ не выполняется, называется дуализацией над произведением частичных порядков.

Важность дуализации обусловлена большим числом приложений, среди которых прежде всего следует выделить логический анализ данных в распознавании (машинное обучение) и поиск ассоциативных правил в базах данных (data mining).

Одним из наиболее востребованных является случай, когда каждое P_i является цепью, т.е. любые два элемента в P_i сравнимы. Если $P_i = \{0, 1\}$ при $i \in \{1, 2, \dots, n\}$ и $0 \prec 1$, то рассматриваемая задача известным образом сводится к построению сокращённой дизъюнктивной нормальной формы монотонной булевой функции от n переменных, заданной конъюнктивной нормальной формой (КНФ) из $|R|$ элементарных дизъюнкций (к дуализации монотонной КНФ).

The problem of dualization over the product of chains: asymptotics for a typical number of solutions

Djukova Elena^{1*}

edjukova@mail.ru

*Masliakov Gleb*²

gleb-mas@mail.ru

*Prokofyev Petr*²

p_prok@mail.ru

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

²Moscow, Lomonosov Moscow State University

³Moscow, Blagonravov Mechanical Engineering Research Institute of RAS

One of the Central problems of logical data analysis is a dualization over the product of partial orders. The problem is formulated as follows.

Let $P = P_1 \times \dots \times P_n$, where P_1, \dots, P_n are finite partially ordered sets. It is believed that the element $y = (y_1, \dots, y_n) \in P$ follows the element $x = (x_1, \dots, x_n) \in P$, if y_i follows x_i for $i = 1, 2, \dots, n$. By $x \prec y$ we denote that $y \in P$ follows $x \in P$ and $x \neq y$. Let $R \subseteq P$, $R^+ = R \cup \{x \in P \mid \exists a \in P, a \prec x\}$. The problem of constructing a 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 $a \prec x$ is not satisfied, is called dualization over the product of partial orders.

The importance of dualization is caused by a large number of applications, among which first of all it is necessary to distinguish logical data analysis in recognition (machine learning) and search of association rules in databases (data mining).

One of the most popular is the case when each P_i is a chain, i.e. any two elements in P_i are comparable. If $P_i = \{0, 1\}$ for $i = 1, 2, \dots, n$ and $0 \prec 1$, the problem under consideration is reduced in a known way to the construction of a reduced disjunctive normal form for a monotonic Boolean function from n variables given by a conjunctive normal form (KNF) of $|R|$ elementary disjunctions (monotonic KNF dualization). An equivalent task is the search for irreducible coverings of Boolean matrix of $|R|$ rows and n columns.

The number of dualization solutions grows almost always exponentially with the problem size (input size). The most efficient algorithm is the one with polynomial step from the input size. However, polynomial algorithms were able to be constructed only for some special cases of monotonic KNF dualization.

Эквивалентной задачей является поиск неприводимых покрытий булевой матрицы из $|R|$ строк и n столбцов.

Число решений дуализации почти всегда растёт экспоненциально с ростом размера задачи (размера входа). Наиболее эффективным считается алгоритм с полиномиальным от размера входа шагом. Однако полиномиальные алгоритмы удалось построить лишь для некоторых частных случаев дуализации монотонной КНФ.

В случае дуализации монотонной КНФ лидерами по скорости счёта являются алгоритмы, базирующиеся на асимптотически оптимальном перечислении решений задачи. Асимптотически оптимальному алгоритму дуализации разрешено делать лишние полиномиальные шаги при условии, что их число почти всегда достаточно мало по сравнению с числом всех решений задачи. Теоретическое обоснование такого алгоритма основано на получении асимптотик типичных значений величины $|I(R)|$ и числа его шагов.

В [1] рассмотрена задача дуализации над произведением цепей в случае, когда $P_i = \{0, 1, \dots, k-1\}$, $k \geq 2$, $i = 1, 2, \dots, n$, и элементы в P_i упорядочены в порядке возрастания. Показано, что исходная задача может быть решена на основе полиномиального перечисления специального подмножества $Q(R)$ наборов из P , содержащего $I(R)$. При условии $|R|^\alpha \leq n \leq d^m$, $\alpha > 1$, $d = k/(k-1)$, $n \rightarrow \infty$, найдены асимптотики типичных значений величин $|I(R)|$ и $|Q(R)|$. Из равенства полученных асимптотик следует, что асимптотически оптимальным является алгоритм дуализации над произведением цепей RUNC-M+, построенный статье: Дюкова Е.В., Масляков Г.О., Прокофьев П.А. О дуализации над произведением частичных порядков // Машинное обучение и анализ данных, 2017. Т. 3. №4. С. 239-249.

Работа поддержана грантом РФФИ № 16-01-00445.

- [1] Дюкова Е. В., Масляков Г. О., Прокофьев П. А. О числе максимальных независимых элементов частичных порядков (случай цепей) // Доклады Академии наук, 2018. Т. 483. №2. (в печати)

In the case of monotonic KNF dualization, algorithms based on asymptotically optimal enumeration of solutions are the leaders in the calculation speed. The asymptotically optimal dualization algorithm is allowed to make extra polynomial steps provided that their number is almost always small enough in compared to the number of all solutions to the problem. The theoretical justification of this algorithm is based on obtaining the asymptotics for the typical values of $|I(R)|$ and the number of it's steps.

In [1] we consider the dualization problem over the product of chains in the case $P_i = \{0, 1, \dots, k - 1\}$, $k \geq 2$, $i = 1, 2, \dots, n$, and the elements in P_i are ordered in ascending order. It is shown that the original problem can be solved on the basis of a polynomial enumeration of a special subset $Q(R)$ of elements of P containing $I(R)$. Asymptotics for typical values of $|I(R)|$ and $|Q(R)|$ are found under the condition $|R|^\alpha \leq n \leq d^m$, $\alpha > 1$, $d = k/(k - 1)$, $n \rightarrow \infty$. From the equality of the obtained asymptotics it follows that the asymptotically optimal is the algorithm for dualization over the product of chains RUNC-M+, built in article: Djukova E. V., Maslyakov G. O., and Prokofyev P. A. 2017. About dualization over the product of partial orders. Machine learning and data analysis. 3(4): 239 – 249.

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

- [1] *Djukova E., Masliakov G., Prokofyev P.* On the number of maximal independent elements of partial orders (the case of chains) // Doklady Mathematics, 2018. 483(2) (in press).

Применение t-распределения Стьюдента при метрической коррекции матриц парных сравнений

*Двоенко Сергей Данилович*¹★

dsd@tsu.tula.ru

*Пшеничный Денис Олегович*¹

denispshenichny@yandex.ru

¹Тула, Тульский государственный университет

DOI: 10.30826/IDP201803

Исходные данные в современных условиях часто представлены взаимными парными сравнениями сходства и различия элементов множества. Данное множество следует погрузить в метрическое пространство для корректного применения методов машинного обучения. Корректное погружение обусловлено неотрицательной определенностью соответствующей матрицы парного сходства элементов.

На практике применяются различные меры сходства и различия. Тем не менее, многие из них не являются корректными метрическими функциями. Поэтому необходимо исправить реальные матрицы парных сравнений так, чтобы гарантировать положительную определенность соответствующих матриц скалярных произведений.

Предложенный подход к восстановлению реальных матриц на основе минимизации отклонений исправленных элементов от их исходных значений ведет к плохо определенным матрицам скалярных произведений с большими числами обусловленности

Развитие предложенного подхода заключается в поиске оптимального числа обусловленности на основе диаграмм чисел обусловленности и отклонений в зависимости от параметра оптимизации. Предлагается проверять статистические гипотезы на основе t-распределения Стьюдента для заданного уровня значимости, т.к. нормированные скалярные произведения можно рассматривать как коэффициенты корреляции.

Работа поддержана грантами РФФИ № 17-07-00319, 15-07-02228, 18-07-00942, 18-07-01087.

- [1] *Dvoenko S.D., Pshenichny D.O.* On metric correction and conditionality of raw featureless data in machine learning // Pattern Recognition and Image Analysis, Pleiades Publishing, 2018. Vol. 28, No.4. (в печати)

Using of Student's t-distribution for metric correction of pairwise comparison matrices

Dvoenko Sergey^{1*}

dsd@tsu.tula.ru

*Pshenichny Denis*¹

denispshenichny@yandex.ru

¹Tula, Tula State University

Modern raw data are usually represented by mutual pairwise comparisons of similarity and dissimilarity of set elements. This set needs to be immersed in a metric space for correct using of machine learning techniques. Correct immersing is subjected to a condition of non-negative definiteness of the corresponding pairwise similarity matrix.

In practice, various similarity and dissimilarity measures are usually used. Nevertheless, many of them are not correct as metric functions. Therefore, it needs to repair real pairwise comparison matrices to guarantee the positive definiteness of corresponding matrices of scalar products.

The proposed approach to repair real matrices based on minimal deviations of their modified elements from initial ones leads to ill-conditioned matrices of scalar products with large condition numbers.

We improve our approach to find optimal condition number based on diagrams of condition numbers and deviations dependent on the optimization parameter. We propose the statistical hypothesis testing based on Student's t-distribution relative to a significance level, since normalized scalar products can be considered as correlation coefficients.

This research is funded by RFBR, grants 17-07-00319, 15-07-02228, 18-07-00942, 18-07-01087.

- [1] *Dvoenko S.D., Pshenichny D.O.* On metric correction and conditionality of raw featureless data in machine learning // Pattern Recognition and Image Analysis, Pleiades Publishing, 2018. Vol. 28, No.4. (in press)

Статистические критерии согласия на основе оценки скользящего экзамена

Неделько Виктор Михайлович

nedelko@math.nsc.ru

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

DOI: 10.30826/IDP201804

В работе исследуются статистические свойства оценки скользящего экзамена как критерия выбора модели решения (метода построения решающей функции).

В задачах анализа данных вместо гипотезы о форме виде распределений мы можем делать предположения непосредственно о виде оптимальной решающей функции. Вид решающей функции в этом случае выступает в роли гипотезы. Такой подход можно считать в некотором смысле обобщением классической теории проверки гипотез. В случае, когда выбор вида решения однозначно связан с выбором вероятностной модели данных, критерий скользящего экзамена фактически является статистическим критерием проверки гипотез.

Для задачи дисперсионного анализа доказано, что критерий скользящего экзамена эквивалентен критерию Фишера для проверки гипотезы однородности при уровне значимости, который соответствует значению статистики Фишера, равному 2. Например, для случая двух подвыборок этот уровень значимости примерно равен 0,16. При этом с ростом числа подвыборок этот уровень значимости стремится к 0.

Установлено также, что применение критерия скользящего экзамена для выбора решающей функции из некоторого однопараметрического класса эквивалентно построению оптимальной решающей функции в рамках байесовской модели с нормальным распределением на значениях параметров.

Работа выполнена при поддержке программы фундаментальных научных исследований СО РАН № I.5.1., проект № 0314-2016-0015, и Российского фонда фундаментальных исследований, проект № 18-07-00600

- [1] *Nedelko V. M.* Statistical Fitting Criterion on the Basis of Cross-Validation Estimation // Pattern Recognition and Image Analysis, ISSN 1054-6618, Pleiades Publishing, Ltd., 2018. Vol. 28, No. 3, — p. 510–515.

Statistical Tests Based on Cross-Validation Estimation

*Victor Nedel'ko**

nedelko@math.nsc.ru

Novosibirsk, Sobolev Institute of Mathematics SB RAS

In this work we investigate the statistical properties of the cross-validation estimate as a statistical criterion for choosing a method for constructing a decision function.

In data analysis we can make assumptions directly about the form of the optimal decision function instead of the hypothesis about the form of distributions. The form of the decision function in this case acts as a hypothesis. Such an approach can be considered as a generalization of the classical theory of hypothesis testing. In some cases, the crossvalidation criterion is in fact a statistical criterion for testing hypotheses.

For the variance analysis problem it is proved that the cross-validation criterion is equivalent to the Fisher's criterion for testing the hypothesis of homogeneity at a significance level that corresponds to the value of Fisher's statistics that is equal to 2. For example, for the case of two subsamples, this significance level is approximately equal to 0,16. At the same time, as the number of subsamples increases, this level of significance tends to 0.

It is also found that the application of the cross-validation criterion for choosing a decision function from a certain one-parameter class is equivalent to constructing an optimal decision function in the framework of the Bayesian model with a normal distribution of the parameter values.

The work was supported by the program of fundamental scientific research of the SB RAS No. I.5.1., Project No. 0314-2016-0015, and the Russian Foundation for Basic Research, grant 18-07-00600.

- [1] *Nedel'ko V. M.* Statistical Fitting Criterion on the Basis of Cross-Validation Estimation // Pattern Recognition and Image Analysis, ISSN 1054-6618, Pleiades Publishing, Ltd., 2018. Vol. 28, No. 3, — p. 510–515.

Исследование свойств распределений байесовских нейронных сетей

Владимирова Мария^{1,2*} `mariia.vladimirova@inria.fr`
*Арбель Джулиан*² `julyan.arbel@inria.fr`

¹Москва, Московский физико-технический институт

²Гренобль, Inria Grenoble Rhône-Alpes

DOI: 10.30826/IDP201805

Байесовский подход позволяет оценить неопределенность предсказания модели и уточнить распределения ее параметров.

Недавними достижениями в нейронных сетях стали модели глубокого обучения, состоящие из большого количества скрытых слоев. Такие модели являются универсальными аппроксиматорами, то есть могут аппроксимировать любую функцию, но изучение их структур не дает никакого интуитивного объяснения. Байесовский подход исследует такие модели, предполагая априорное распределение на их параметрах и вычисляет апостериорное распределение при имеющихся реальных данных. В последние годы было предложено множество вариационных алгоритмов, позволяющих эффективно реализовать байесовский вывод для глубоких архитектур.

Данная работа посвящена исследованию свойств априорных распределений скрытых нейронов в байесовских нейронных сетях. Проводится асимптотическое сравнение распределений нейронов с субвейбулловским, оценивающим тяжесть хвоста распределения. При определенных условиях на функцию активации распределение скрытых нейронов в ℓ -м слое является субвейбулловским с параметром равным $\ell/2$. Это означает, что хвосты распределений скрытых нейронов в таких нейронных сетях становятся тяжелее при увеличении глубины модели. Показано, что ковариационная матрица с увеличением числа слоев становится более разрежена.

Работа поддержана исследовательским центром Inria Grenoble Rhône-Alpes.

- [1] *Vladimirova M., Arbel J., Mesejo P.* Bayesian neural networks become heavier-tailed with depth // Preprint, 2018.

Distributional properties of Bayesian neural networks

Vladimirova Mariia^{1,2*}

mariia.vladimirova@inria.fr

*Arbel Julyan*²

julyan.arbel@inria.fr

¹Moscow, Moscow Institute of Physics and Technology

²Grenoble, Inria Grenoble Rhône-Alpes

Bayesian approach infers distributions in machine learning models. It estimates the uncertainty of model predicts.

Recent advances in neural networks are the deep models which include enormous number of layers. However, the deep models are not interpretable in the sense that while it can approximate any function, studying its structure will not give any intuitive explanation. Bayesian approach investigates such models assuming a prior distribution on the weights and computes the posterior distribution observing real data. In the last years, variational inference algorithms was proposed to make the Bayesian framework implementation applicable for deep architectures.

This study is devoted to the investigation of hidden units prior distributions in Bayesian neural networks. We extend the theoretical understanding of feedforward fully connected neural networks. The weights are assumed to be independent and normally distributed. We define a sub-Weibull distributional family whose parameter characterizes the tails of the distributions. Under certain condition for activation functions, the hidden unit distribution in the ℓ -th layer is sub-Weibull with tail parameter equals to $\ell/2$. It means that hidden layer units in such neural networks follow heavier-tailed distributions depending on increasing number of layer. It leads to increasing sparsity in hidden units with increasing depth.

This research is funded by Inria Grenoble Rhône-Alpes.

- [1] *Vladimirova M., Arbel J., Mesejo P.* Bayesian neural networks become heavier-tailed with depth // Preprint, 2018.

Новый подход к специфичности возможностей мер и его роль в теории принятия решений

Зубюк Андрей Владимирович¹

zubjuk@physics.msu.ru

¹Москва, МГУ имени М. В. Ломоносова, физический факультет

DOI: 10.30826/IDP201806

В работе исследуется проблема определения отношения специфичности в теории возможностей Пытьева, в которой возможность — это монотонная мера Π , числовые значения которой, кроме 0 и 1, не имеют содержательной интерпретации. Содержательно интерпретированы могут быть только следующие утверждения и их следствия: $\Pi(A) = 0$ (A невозможно), $\Pi(A) > \Pi(B)$ (A более возможно, чем B), $\Pi(A) = \Pi(B)$ (A и B равновозможны). Любые выводы теории возможностей Пытьева инвариантны относительно строго монотонных полунепрерывных снизу преобразований значений возможностей с неподвижными точками 0 и 1. Это позволяет эксперту задавать числовые значения возможностей в своей собственной субъективной порядковой шкале.

В теории возможностей Заде, основанной на нечётких множествах, специфичность распределений возможностей определяется посредством отношения включения нечётких множеств: распределение π_1 не менее специфично, чем π_2 , если поточечно $\pi_1(\cdot) \leq \pi_2(\cdot)$. Однако, в теории возможностей Пытьева такой подход неприменим, если числовые значения распределений π_1 и π_2 заданы в разных субъективных шкалах (разными экспертами). Согласно основной идее работы отношение специфичности должно быть согласовано с возможностными методами принятия решений: чем специфичнее возможностная модель, тем уже должно быть множество оптимальных решений (решение оптимально, если минимизирует возможность ошибки). В работе определено и исследовано такое отношение специфичности. Полученные результаты применимы не только в рамках теории возможностей Пытьева, но и в рамках сравнительной теории возможностей, качественной теории принятия решений и т. д. Работа поддержана грантами РФФИ №№ 17-07-00832, 18-07-00424.

- [1] Zubjuk A. V. A new approach to specificity in possibility theory: Decision-making point of view // Fuzzy Sets and Systems, Elsevier, 2018. (в печати). <https://doi.org/10.1016/j.fss.2018.06.017>.

A new approach to specificity in possibility theory: Decision-making point of view

Zubyuk Andrey¹

zubjuk@physics.msu.ru

¹Moscow, M. V. Lomonosov Moscow State University, Faculty of Physics

The paper investigates the problem of defining a specificity relation in Pyt'ev possibility theory. That branch of possibility theory developed by prof. Yu. P. Pyt'ev in the late 1990s defines possibility as a monotone measure Π which numerical values except 0 and 1 are meaningless. The following information and its consequent implications are meaningful: $\Pi(A) = 0$ (A is impossible), or $\Pi(A) > \Pi(B)$ (A is more possible than B), or $\Pi(A) = \Pi(B)$ (A and B are equally possible). All results obtained with Pyt'ev possibility theory are invariant to any strictly increasing lower semi-continuous transformation of all possibility values with the fixed points: 0 and 1. This fact allows anyone to assign numerical values to possibilities using an individual subjective ordinal scale of possibility values.

Unlike Zadeh possibility theory based on fuzzy sets, where specificity of possibility distributions is evaluated by means of fuzzy set inclusion, and distribution π_1 is more specific than π_2 iff $\pi_1(\cdot) \leq \pi_2(\cdot)$ pointwise, in Pyt'ev possibility theory such an approach can not be applied if numerical values of π_1 and π_2 are assigned through different subjective scales (do not commensurate). A thesis we put forward in this paper states that the specificity relation definition has to be consistent with the possibilistic decision-making approaches. The following is intended to be true: the more specific an underlying possibility and a corresponding possibility distribution are, the narrower a set of the optimal decisions is (a decision is optimal if it minimizes the possibility of error). A specificity relation satisfying such a requirement is defined and studied. The obtained results are not limited to Pyt'ev possibility theory. They can be applied to comparative possibility orderings, qualitative decision making, etc. as well. This research is funded by RFBR, grants 17-07-00832, 18-07-00424.

- [1] Zubyuk A. A new approach to specificity in possibility theory: Decision-making point of view // Fuzzy Sets and Systems, Elsevier, 2018. (in press). <https://doi.org/10.1016/j.fss.2018.06.017>.

Математический формализм субъективного моделирования

*Пытьев Юрий Петрович*¹

yuri.pytyev@gmail.com

*Фаломкина Олеся Владимировна*¹

olesya.falomkina@gmail.com

*Чуличков Алексей Иванович*¹

achulichkov@gmail.com

*Шишкин Семен Александрович*¹

darkshish@gmail.com

¹Москва, МГУ им. М. В. Ломоносова

DOI: 10.30826/IDP201807

Рассмотрен математический формализм субъективного моделирования (МФСМ) в задаче восстановления модели измерений. При её решении использован научный опыт модельера-исследователя (м-и) и его субъективные суждения о модели. Модели датчика температуры и погрешности измерений не заданы.

МФСМ создан для моделирования недостоверности субъективных знаний, нечеткости их содержания и неполноты, начиная с “абсолютного незнания” вплоть до “точного знания” объекта исследования (ОИ). Основа МФСМ – пространство $(X, \mathcal{P}(X), \text{Pl}^{\tilde{x}}, \text{Bel}^{\tilde{x}})$ с мерами правдоподобия $\text{Pl}^{\tilde{x}}(\cdot): \mathcal{P}(X) \rightarrow \mathcal{L}$ и доверия $\text{Bel}^{\tilde{x}}(\cdot): \mathcal{P}(X) \rightarrow \hat{\mathcal{L}}$, $\mathcal{L}(\hat{\mathcal{L}}) = ([0, 1], \leq (\geq), +(\hat{+}), \times(\hat{\times}))$, сложения: $a + b = \max\{a, b\}$, $a \hat{+} b = \min\{a, b\}$, умножения: $a \times b = \min\{a, b\}$, $a \hat{\times} b = \max\{a, b\}$, $a, b \in [0, 1]$, X – множество значений неизвестного параметра x модели $M(x)$ ОИ, $\mathcal{P}(X)$ – все подмножества X , неопределенный элемент \tilde{x} со значениями в X характеризует модальности субъективных суждений м-и об истинности $x \in X$ правдоподобием $\text{Pl}^{\tilde{x}}(\tilde{x} = x)$ и доверием $\text{Bel}^{\tilde{x}}(\tilde{x} \neq x)$. Меры $\text{Pl}^{\tilde{x}}$ и $\text{Bel}^{\tilde{x}}$ любого $E \in \mathcal{P}(X)$ суть $\text{Pl}^{\tilde{x}}(E) = \text{Pl}^{\tilde{x}}(\tilde{x} \in E) = \sup_{x \in E} t^{\tilde{x}}(x)$, $E \neq \emptyset$, $\text{Pl}^{\tilde{x}}(\emptyset) \stackrel{\text{def}}{=} 0$; $\text{Bel}^{\tilde{x}}(E) = \text{Bel}^{\tilde{x}}(\tilde{x} \in E) = \inf_{x \in X \setminus E} \hat{t}^{\tilde{x}}(x)$, $E \neq X$, $\text{Bel}^{\tilde{x}}(X) \stackrel{\text{def}}{=} 1$, где $t^{\tilde{x}}(x) = \text{Pl}^{\tilde{x}}(\tilde{x} = x)$, $\hat{t}^{\tilde{x}}(x) = \text{Bel}^{\tilde{x}}(\tilde{x} \neq x)$, $x \in X$, суть распределения правдоподобий и доверий, заданные м-и.

Работа поддержана грантами РФФИ № 17-07-00832_а и № 18-07-00424_а.

[1] Пытьев Ю. П., Фаломкина О. В., Шишкин С. А., Чуличков А. И. Математический формализм субъективного моделирования // Машинное обучение и анализ данных, Москва, 2018.

Mathematical formalism of subjective modeling

*Pytyev Yuri*¹

yuri.pytyev@gmail.com

*Falomkina Olesya*¹

olesya.falomkina@gmail.com

Chulichkov Alexey^{1*}

achulichkov@gmail.com

*Shishkin Semyon*¹

darkshish@gmail.com

¹Russia, Moscow, M. V. Lomonosov Moscow State University

The mathematical formalism for subjective modeling (MFSM) within the problem of reconstruction of the measurement model is considered. To solve it the scientific experience of the researcher-modeler (r-m) and his subjective judgments are being used. The temperature and measurement error mathematical models are unknown.

The MFSM was created for the modeling of the subjective knowledge unreliability, its matter fuzziness and incompleteness ranging from the “absolute ignorance” up to the “complete knowledge” of the model of the research object (RO). The MFSM is based on the space $(X, \mathcal{P}(X), \text{Pl}^{\tilde{x}}, \text{Bel}^{\tilde{x}})$, where $\text{Pl}^{\tilde{x}}(\cdot): \mathcal{P}(X) \rightarrow \mathcal{L}$ and $\text{Bel}^{\tilde{x}}(\cdot): \mathcal{P}(X) \rightarrow \hat{\mathcal{L}}$ are the *plausibility* and *belief measures*, $\mathcal{L}(\hat{\mathcal{L}}) = ([0, 1], \leq (\geq), +(\hat{+}), \times(\hat{\times}))$, the additions are $a + b = \max\{a, b\}$, $a \hat{+} b = \min\{a, b\}$, the multiplications are $a \times b = \min\{a, b\}$, $a \hat{\times} b = \max\{a, b\}$, $a, b \in [0, 1]$, X is the set of all possible values of the unknown parameter x of the RO model $M(x)$, $\mathcal{P}(X)$ is the set of all subsets of X , the *indeterminate element* $\tilde{x} \in X$ characterizes the *subjective judgments modalities made by the r-m about the validity of $x \in X$ by measures $\text{Pl}^{\tilde{x}}(\tilde{x} = x)$ and $\text{Bel}^{\tilde{x}}(\tilde{x} \neq x)$* . For any $E \in \mathcal{P}(X)$ $\text{Pl}^{\tilde{x}}$ and $\text{Bel}^{\tilde{x}}$ are given by formulas $\text{Pl}^{\tilde{x}}(E) = \text{Pl}^{\tilde{x}}(\tilde{x} \in E) = \sup_{x \in E} t^{\tilde{x}}(x)$, $E \neq \emptyset$, $\text{Pl}^{\tilde{x}}(\emptyset) \stackrel{\text{def}}{=} 0$; $\text{Bel}^{\tilde{x}}(E) = \text{Bel}^{\tilde{x}}(\tilde{x} \in E) = \inf_{x \in X \setminus E} \hat{t}^{\tilde{x}}(x)$, $E \neq X$, $\text{Bel}^{\tilde{x}}(X) \stackrel{\text{def}}{=} 1$, where $t^{\tilde{x}}(x) = \text{Pl}^{\tilde{x}}(\tilde{x} = x)$, $\hat{t}^{\tilde{x}}(x) = \text{Bel}^{\tilde{x}}(\tilde{x} \neq x)$, $x \in X$, are *plausibility and belief distributions* defined by the r-m.

This research is funded by RFBR, grants 17-07-00832-a and 18-07-00424.a.

- [1] *Pytyev Yu, Falomkina O., Shishkin S., Chulichkov A.* Mathematical formalism of subjective modeling // Machine Learning and Data Analysis, Moscow, 2018.

Информативные априорные предположения в задаче привилегированного обучения

Нейчев Радослав Георгиев¹*

neychev@phystech.edu

Стрижов Вадим Викторович¹

strijov@gmail.com

¹Москва, Лаборатория машинного интеллекта МФТИ

DOI: 10.30826/IDP201808

Решается задача выбора прогностической модели оптимального правдоподобия. Для снижения вычислительной сложности модели используется смесь экспертов — ансамбль в котором в качестве локальных моделей-экспертов выступают интерпретируемые модели. Для интерпретируемой модели известно, как изменение параметров модели влияет на точность и устойчивость аппроксимации. Для повышения правдоподобия привлекается априорная привилегированная информация. В качестве этой информации выступает априорная разметка. В данной работе этой разметкой является априорное знание о принадлежности объекта-области изображения к некоторому классу в задаче определения границы радужки. Вводятся две модели: модель-учитель и модель-ученик. Модель-учитель превосходит ученика в сложности и точности, и ее предсказания используются в качестве априорной привилегированной информации при построении модели-ученика, которую требуется оптимизировать. Использование априорной привилегированной информации о решаемой задаче ускоряет сходимость оптимизации параметров модели, повышает качество ее предсказаний и снижает ее сложность без значимых потерь в качестве. Данный подход проиллюстрирован задачей определения границы радужки на фотографиях глаз.

Работа выполнена при поддержке РФФИ № 16-07-01163.

- [1] *Бахтеев О. Ю., Стрижов В. В.* Deep Learning Model Selection of Suboptimal Complexity // Autom Remote Control, volume 79, Springer, 2018. — p.1474–1488. <https://link.springer.com/article/10.1134/S000511791808009X>.

Informative prior in privileged learning problems

*Radoslav Neychev*¹*

neychev@phystech.edu

*Vadim Strijov*¹

strijov@gmail.com

¹Moscow, MIPT Machine Intelligence Laboratory

The problem of model selection of optimal likelihood is considered. The ensemble of several interpretable models is used to both boost the likelihood and reduce the computational complexity. Interpretable model retains the interconnection between its robustness, accuracy and parameters. External prior data markup is referred as informative prior. Iris boundary localization problem serves as inspiration and for testing purpose. Teacher-model and student-model are introduced. Teacher-model is more complex and reliable. Its predictions are used as privileged prior for student-model. Usage of privileged information allows to speed up the convergence of student's parameters optimization and to boost the predictions quality of the less complex model.

This research was supported by the RFBR project 16-07-01163.

- [1] *O. Yu. Bakhteev, V V. Strijov* Deep Learning Model Selection of Sub-optimal Complexity // Autom Remote Control, volume 79, Springer, 2018. — p. 1474–1488. <https://link.springer.com/article/10.1134/S000511791808009X>.

Байесовский выбор наиболее правдоподобной структуры модели глубокого обучения

*Бахтеев Олег Юрьевич*¹

bakhteev@phystech.edu

Стрижов Вадим Викторович^{1,2}

strijov@ccas.ru

¹Москва, Московский физико-технический институт

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

DOI: 10.30826/IDP201809

В работе рассматривается задача выбора структуры модели глубокого обучения. Модель — это вычислительный вероятностный граф, т.е. граф, в котором ребрами выступают нелинейные функции, а вершинами — результаты действия функцией на выборку. Каждому ребру поставлено в соответствие множество нелинейных функций, такое что линейная комбинация этих функций определяет дифференцируемую функцию заданной сигнатуры. Структурой модели назовем веса линейной комбинации этих функций. Для нахождения оптимальной структуры предлагается ввести вероятностную интерпретацию модели, т.е. предположения о распределениях параметров и структуры модели. Проводится градиентная оптимизация параметров и гиперпараметров модели на основе байесовского вариационного вывода. Решается двухуровневая задача оптимизации: на первом уровне проводится оптимизация нижней оценки правдоподобия модели по вариационным параметрам модели. На втором уровне проводится оптимизация гиперпараметров модели. В качестве оптимизируемой функции для гиперпараметров модели предлагается обобщенная функция правдоподобия. Показано, что данная функция позволяет проводить оптимизацию несколькими алгоритмами: последовательным добавлением и удалением параметров, полным перебором, а также максимизацией нижней оценки правдоподобия модели. Проводится сравнение с эвристическими алгоритмами выбора структуры модели. Вычислительный эксперимент проводится на синтетических данных и выборке рукописных цифр MNIST.

Работа поддержана грантом РФФИ № 16-07-01154.

- [1] *Бахтеев О. Ю., Стрижов В. В.* Выбор модели глубокого обучения субоптимальной сложности // Автоматика и телемеханика, Москва, 2018. —№. 8, С. 129–147.

Bayesian deep learning optimal model structure selection

*Oleg Bakhteev*¹

bakhteev@phystech.edu

Vadim Strijov^{1,2}

strijov@ccas.ru

¹Moscow, Moscow Institute of Physics and Technology

²Moscow, FRCCSC of the Russian Academy of Sciences

The paper presents a method of deep learning model structure selection. The model is a computation graph. Its edges correspond to some nonlinear functions. Its vertices correspond to the intermediate variables. Each edge refers to a set of primitive functions. A linear combination of these functions is a differential function of some given signature. The structure of the model represents a vector of weights for linear combinations of primitive functions. The author proposes a probabilistic interpretation of the model. The model parameters and hyperparameters are optimized using gradient methods. The optimization of the model is based on the Bayesian variational inference. The optimization is a bi-level optimization problem. The first level corresponds to the evidence lower bound optimization with regard to variational model parameters. The second level corresponds to the hyperparameter optimization. The author proposes a generalized loss function for this level. This function allows to optimize the hyperparameters with different algorithms: sequential model growth and pruning, bruteforce optimization and evidence lower bound optimization. The proposed method is compared with heuristic model structure selection methods. The author performs computational experiment on a synthetic dataset and on a dataset of handwritten digits dataset MNIST.

- [1] *O. Yu. Bakhteev and V. V. Strijov* Deep Learning Model Selection of Suboptimal Complexity // Automation and Remote Control, Moscow, 2018, Vol. 79, No. 8, pp.1474–1488.

GLCM, kNN and Meanshift в задаче детектирования нейронов по изображениям срезов мозга, окрашенных по Нисслию

Носова Светлана Александровна* svetlana.nosova@itmm.unn.ru
Турлапов Вадим Евгеньевич vladim.turlapov@gmail.com

¹Нижний Новгород, ННГУ им. Н. И. Лобавчевского

DOI: 10.30826/IDP201810

Разработан метод обнаружения нейронов на изображениях срезов мозга, окрашенных по Нисслию. Метод использует текстурные признаки нейронов, построенные на основе 4x матриц взаимной встречаемости (GLCM). Метод включает в себя следующие этапы: предобработка изображений, классификация пикселей по текстурным признакам алгоритмом kNN и кластеризация пикселей нейронов алгоритмом Meanshift. Предобработка включает в себя следующие шаги: конвертация в оттенки серого, выравнивание гистограммы, квантование гистограммы. Применены и изучены различные способы преобразования цветного изображения в оттенки серого. Наилучший результат дает преобразование по синей компоненте цвета. Показано, что использование квантования гистограммы на 2 и 4 бина дает близкое качество детектирования с квантованием на 8 бин ($F1 = 0,83..0,85$). Результаты показывают, что kNN является лучшим выбором для текущей задачи классификации по сравнению с NBC. Наш алгоритм обеспечивает следующее качество детектирования: $precision = 0,82$; $recall = 0,92$; $F1 = 0,86$. Предложенный метод показал лучший результат по сравнению с аналогами. Планируется продолжить исследования на расширенном наборе данных и данных с социальнозначимыми заболеваниями мозга.

- [1] Носова С. А., Турлапов В. Е. GLCM, kNN and Meanshift в задаче детектирования нейронов по изображениям срезов мозга, окрашенных по Нисслию // Машинное обучение и анализ данных, Москва: Вычислительный центр им. А.А. Дородницына РАН, 2018.

GLCM, kNN and Meanshift for neuron detection on Nissl-stained brain slice image

*Svetlana Alexandrovna**

svetlana.nosova@itmm.unn.ru

Vadim Evgenievich

vadim.turlaov@gmail.com

¹City, Institution

²City, Institution

The method for neuron detection on Nissl-stained brain slice images is proposed. The method uses textural features of neurons extracted from 4 GLC-matrices. The method includes the following steps: image preprocessing, kNN classification by the textural features and Meanshift clustering of neuron pixels. Preprocessing includes the following steps: grayscale conversion, histogram equalization, histogram quantization. Gray conversion by blue component gives the best result. It is shown that using 2-,4- bin histogram gives close detection quality with 8-bin histogram ($F1 = 0,83..0,85$). For pixel classification kNN algorithm was used. The results demonstrate that kNN is better choice for current task in comparing with NBC. The reached detection quality for given approach is $precision = 0,82, recall = 0,92, F1 = 0,86$. It is shown that our results are near the same or some better in *recall* characteristic in comparing with other neuron detection method. In our future work we'll prolong this investigation for great volume of dataset and special dataset for important diseases.

- [1] *Nosova S. A, Turlapov V. E* GLCM, kNN and Meanshift for neuron detection on Nissl-stained brain slice images // Machine Learning and Data Analysis, Moscow: Computing Centre of. A.A. Dorodnicyn RAS, 2018.

Весовой случайный лес для построения управляемых моделей выживаемости

Уткин Лев Владимирович^{1*}

lev.utkin@gmail.com

Мелдо Анна Александровна^{1,2}

anna.meldo@yandex.ru

¹Санкт-Петербург, Санкт-Петербургский политехнический университет Петра Великого

²Санкт-Петербург, Санкт-Петербургский клинический онкоцентр

DOI: 10.30826/IDP201811

Большое число моделей выживаемости было разработано в последние десятилетия для обобщения таких моделей, как Каплана-Майера и Кокса. Это – модели на основе метода Лассо, SVM, нейронные сети. Для работы с ограниченными данными, случайные леса выживаемости (СЛВ) как обобщение обычных случайных лесов были представлены многими авторами и показали лучшие результаты для многих наборов данных. В работе предлагается взвешенный СЛВ как модификация стандартного СЛВ. В соответствии с ней, усреднение функций риска на выходе каждого дерева решений, которое используется для вычисления функции риска всего леса, заменяется взвешенной суммой этих функций, т.е. веса присваиваются деревьям и рассматриваются как параметры обучения, вычисляемые решением задачи оптимизации, максимизирующей индекс конкордации или C-индекс. Задача оптимизации для вычисления весов сведена к квадратичной задаче с линейными ограничениями заменой индикаторных функций в C-индексе петлевыми функциями потерь. Основной причиной введения весов деревьев является управление результатами анализа выживаемости путем изменения функции потерь в соответствии с задачами приложения так же, как делается в нейронных сетях. Другая причина – использование взвешенных СЛВ как элементов каскадной архитектуры, такой как глубокий лес. Численные эксперименты показывают преимущество предлагаемой модели по сравнению с обычным RSF для многих наборов данных. Исследование выполнено за счет гранта Российского научного фонда (проект № 18-11-00078).

- [1] *Utkin L. V., Meldo A. A.* A deep survival forest for constructing controllable survival models // Knowledge-Based Systems, 2018 (в печати).

A weighted random survival forest for constructing controllable models

*Utkin Lev*¹★

lev.utkin@gmail.com

Meldo Anna^{1,2}

anna.meldo@yandex.ru

¹St.Petersburg, Peter the Great St.Petersburg Polytechnic University

²St.Petersburg, St.Petersburg Clinical Oncology Center

A lot of survival models have been developed last decades to solve many application problems and to generalize well-known models such as the Kaplan-Meier estimator, the Cox model. These are the Lasso-based Cox models, survival SVMs, neural networks. In order to deal with limited survival data, random survival forests (RSF) as an extension of original random forests have been presented by many authors and demonstrated outperforming results for many data sets. A weighted RSF as a modification of the original RSF is proposed in the paper. According to the modification, the standard averaging of cumulative hazard functions produced by every decision tree, which is used for computing the forest cumulative hazard function, is replaced by the weighted sum of these functions. In fact, the weights are assigned to decision trees, and they do not depend on the training examples. The weights are regarded as training parameters which can be computed in an optimal way by solving an optimization problem that maximizes the concordance error rate called the C-index.

It is shown that the optimization problem for computing optimal weights can be reduced to the standard quadratic optimization problem with linear constraints if we replace indicator functions in the C-index with the well-known hinge loss functions. The main reason for introducing the weights of trees is to control the survival analysis results by changing the loss function in accordance with the application tasks like neural networks. Another reason is to view the weighted RSFs as components of a more complex cascade architecture like the deep forest. Numerical experiments illustrate the weighted RSF and show that the proposed model may outperform the original RSF.

This work is supported by the Russian Science Foundation under grant 18-11-00078.

- [1] *Utkin L. V., Meldo A. A.* A deep survival forest for constructing controllable survival models // Knowledge-Based Systems, 2018 (in press).

Автоматизированная система дифференциальной диагностики заболеваний лёгких

Мелдо Анна Александровна^{1,2} *

anna.meldo@yandex.ru

*Уткин Лев Владимирович*¹

lev.utkin@gmail.com

¹Санкт-Петербург, Санкт-Петербургский политехнический университет Петра Великого

²Санкт-Петербург, Санкт-Петербургский клинический онкоцентр

DOI: 10.30826/IDP201812

Несмотря на разнообразие систем диагностики рака легкого, большинство из них обучены на узловой его форме и не эффективны в распознавании других форм рака и дифференциальной диагностике со схожими по КТ-картине заболеваниями. По данным Санкт-Петербургского Онкоцентра лишь 65% больных раком лёгкого имели типичную картину узлового рака, в 26% случаев изображения КТ соответствовали нескольким заболеваниям, что требовало дополнительных дифференциально-диагностических критериев, в 9% случаев рак лёгкого был крайне трудно распознаваем по КТ в связи с атипичной картиной. Для решения проблемы обнаружения рака предлагаются четыре подхода. Первый — применение гистограмм длин хорд для представления поверхности образования. Второй – применение гистограмм плотностей, построенных по точкам на хордах с учетом значений Хаунсфилда в этих точках. Третий подход – создание базы данных, который содержит “экзотические” случаи образований. Четвертый – использование гистограмм для обучения сиамской нейронной сети и сиамского глубокого леса, которые являются основой предлагаемой архитектуры системы диагностики и позволяют принимать решение аналогично тому, как это делает врач. Сиамская нейронная сеть обучается на основе всех пар образований из созданной базы. В режиме использования семантически близкий пример из базы соответствует диагнозу для нового образования. Такая архитектура копирует выводы врача и позволяет интерпретировать диагноз. Исследование выполнено за счет гранта Российского научного фонда (проект № 18-11-00078).

- [1] *Meldo A. A., Utkin L. V.* A computer-aided system for differential diagnosis of lung diseases // *Artificial Intelligence in Medicine*, 2018 (в печати).

A computer-aided system for differential diagnosis of lung diseases

Meldo Anna^{1,2}★

anna.meldo@yandex.ru

*Utkin Lev*¹

lev.utkin@gmail.com

¹St.Petersburg, Peter the Great St.Petersburg Polytechnic University

²St.Petersburg, St.Petersburg Clinical Oncology Center

In spite of a wide variety of computer-aided systems (CAS) for diagnosing lung cancer, most of them are trained by using the nodule cancer form and may be not effective when we deal with different forms of cancer and with the differential diagnosis with similar CT images (tuberculosis, aspergilloma, hamartoma, etc.). According to the data obtained in the St.Petersburg Oncology Center, only 65% of patients with lung cancer have a typical nodular cancer representation, 26% of patients have different diseases. This requires additional differential diagnostic criteria. Moreover, 9% of patients have the lung cancer which is extremely difficulty recognized due to its atypical visualization. To solve the problem of detecting all types of lung cancer, it is proposed four ideas. The first idea is to apply the length chord histograms for representing the shape peculiarities of a tissue. The second idea is to apply the density value histogram constructed on the basis of points on the chords and the Hounsfield unit values at these points. The third idea is to collect and to use a special dataset which contains a lot of “exotic” cases of cancer and non-cancer tissues. The fourth idea is to use the obtained histogram representation for training a Siamese neural network and the Siamese Deep Forest which can be regarded as a key element of the proposed CAS architecture and allows us to make decision in a way similar to the doctor’s decision making. The Siamese network is trained by using all possible tissue pairs from the “exotic” dataset. This architecture “copies” the doctor’s reasoning and allows us to interpret the diagnosis. This work is supported by the Russian Science Foundation under grant 18-11-00078.

- [1] *Meldo A. A., Utkin L. V.* A computer-aided system for differential diagnosis of lung diseases // Artificial Intelligence in Medicine, 2018 (in press).

Определение релевантности параметров нейросети методом Белсли

*Грабовой Андрей Валериевич*¹* grabovoy.av@phystech.edu
*Бахтеев Олег Юриевич*¹ bakhteev@phystech.edu
Стрижов Вадим Викторович^{1,2} strijov@ccas.ru

¹Москва, Московский физико-технический институт

²Москва, Вычислительного центра им. А.А. Дородницына

ФИЦ ИУ РАН

DOI: 10.30826/IDP201813

Работа посвящена оптимизации структуры нейронной сети. Число параметров нейросети можно существенно снизить без значимой потери качества и значимого повышения дисперсии функции ошибки. Предлагается метод прореживания параметров нейронной сети, основанный на автоматическом определении релевантности параметров. Для определения релевантности параметров, анализировалась ковариационная матрица апостериорного распределения параметров и из нейросети удалялись наименее релевантные и мультиколлинеарные параметры. Для определения мультиколлинеарности использовался метод Белсли.

В качестве базовых моделей использовались метод оптимального прореживания и метод основанный на вариационном выводе. В работе рассматривается новый метод прореживания параметров. Предложенный метод учитывает зависимость между параметрами нейросети внутри каждого слоя и удаляет те параметры, которые являются максимально зависимы между собой.

Для анализа качества представленного алгоритма проводились эксперименты на выборке Boston Housing, а также на синтетических данных. В ходе эксперимента было получено, что нейросети прорежены методом Белсли являются более устойчивы к шуму на входных данных. Также было показано, что качество прогноза нейросетей после удаления значительного числа параметров не ухудшается.

Работа поддержана грантом РФФИ № 16-07-01154.

- [1] *Грабовой А.В. Бахтеев О.Ю. Стрижов В.В.* Автоматическое определение релевантности параметров нейросети // Информатика и ее применения, 2019 (в печати).

Assessment of the relevance of the Neural network parameters with Belsly method

*Grabovoy Andrey*¹★

grabovoy.av@phystech.edu

*Bakhteev Oleg*¹

bakhteev@phystech.edu

Strijov Vadim^{1,2}

strijov@ccas.ru

¹Moscow, Moscow Institute of Physics and Technology

²Moscow, FRC CSC RAS

This paper investigates methods of optimisation neural network structure. A number of neural network parameters can be reduced without significant loss of quality. The paper proposes a method for automatically determining the relevance of parameters to prune the neural network. This method analyses the covariance matrix of the posterior distribution of the model parameters and removes the least relevant and multicollinear parameters. It uses the Belsly method to search the multicollinearity in the neural network.

Optimal Brain Damage and Practical Variational Inference were used as basic methods. The paper proposes a new method to prune a neural network. This method is looking for the most dependent parameters in each layer of a neural network and removes them.

The proposed method was tested on the Boston Housing dataset and synthetic data. The experiments shown that neural networks, which were pruning by the Belsly method, are more resistant to noise in input data. It was also shown that the quality after removing a significant number of the neural network parameters does not deteriorate.

The research was supported by the RFBR, project 16-07-01154.

- [1] *Grabovoy A. Bakhteev O. Strijov V.* Automatic search for the relevance of neural network parameters // Informatics and Applications, 2019 (in press).

Быстрое приближенное обучение двухклассовому распознаванию по методу опорных векторов в условиях большого числа объектов

*Макарова Александра Игоревна**

aleksarova@gmail.com

Сулимова Валентина Вячеславовна

vsulimova@yandex.ru

Россия, Тула, ТулГУ

DOI: 10.30826/IDP201814

Одним из наиболее удобных и надежных инструментов решения задачи обучения двухклассовому распознаванию является метод опорных векторов (Support Vector Machines, SVM). Особенностью современных прикладных задач является необходимость обучения в условиях большого числа объектов. При этом процесс обучения оказывается весьма трудоемким и требует повышения производительности существующих реализаций методов обучения и разработки новых. В данной работе предлагается простой подход, основанный на усреднении решающих правил SVM, построенных в линейном признаковом пространстве по небольшим подвыборкам объектов, случайным образом выбранным из обучающей совокупности. Экспериментальные исследования на модельных и реальных данных показали, что данный подход позволяет в условиях большого числа объектов достаточно быстро найти приближенное, но не сильно отличающееся от точного решение задачи SVM. Более того, в ряде случаев получаемое решение характеризуется меньшим процентом ошибок на тестовой совокупности по сравнению с результатом, получаемым при помощи традиционных подходов, в частности при помощи библиотеки LibSVM. Кроме того, предложенный подход является экономичным по памяти, обеспечивая возможность его применения для обучения на одной вычислительной машине, и, в то же время, обладает высокой степенью параллелизма по данным и может быть эффективно реализован с применением технологий параллельных и распределенных вычислений. Работа выполнена при поддержке РФФИ, грант № 18-07-01087.

- [1] Макарова А. И., Сулимова В. В. Быстрое приближенное обучение двухклассовому распознаванию по методу опорных векторов в условиях большого числа объектов // Машинное обучение и анализ данных, 2018 (в печати).

Fast approximate two-class SVM learning in the case of a large number of objects

*Makarova Alexandra**

aleksarova@gmail.com

Sulimova Valentina

vsulimova@yandex.ru

Tula, Russia, TulSu

One of the most convenient and reliable tools for solving the two-class recognition problem is Support Vector Machines (SVM). A feature of modern applied recognition problems is the need for learning in a large number of objects. At that the learning process is very laborious and requires increasing the productivity of existing realizations of learning methods and developing new ones.

In this paper we propose a simple approach based on averaging the SVM decision rules, which are constructed in a linear feature space using small subsamples of objects randomly selected from the training set.

Experimental research on model and real data have shown that in the case of a large number of objects this approach allows to quickly find an approximate solution of the SVM problem that is not very different from the exact one. Moreover, in a number of cases it allows to reach a lower error rate in the test set in contrast to results obtained using traditional instruments, such as LibSVM library.

In addition, the proposed approach is memory-efficient and so provides the possibility training on a single computer, and, at the same time, has a high degree of parallelism in data and can be effectively implemented using parallel and distributed computing technologies.

This research is funded by RFBR, grant 18-07-01087.

- [1] *Makarova A. I., Sulimova V. V.* Fast approximate two-class SVM learning in the case of a large number of objects // Machine Learning and Data Analysis, 2018 (in press).

Метод быстрого множественного попарного выравнивания на основе префиксных деревьев

Яковлев Павел Андреевич¹*

yakovlev@biocad.ru

¹Санкт-Петербург, ЗАО БИОКАД

DOI: 10.30826/IDP201815

Моноклональные антитела уже более двадцати лет являются основной для создания биологических лекарственных средств против онкологических и аутоиммунных заболеваний. Высокая эффективность такой терапии связана с уникальными свойствами подобных молекул, включающими в себя как высокую изменчивость, обеспечивающую возможность получить молекулы, связывающиеся с каждым необходимым белком, так и крайнюю консервативность организации их последовательности. Неугасающий интерес к антителам подстегивает использование новейших методов их анализа, в том числе высокопроизводительное секвенирование, способное получать миллионы последовательностей различных антител.

Анализ получаемых данных включает в себя аннотацию – отнесения каждой прочитанной последовательности к одному из десятков тысяч классов, а также выделение на последовательности особых регионов. Вычислительно это требует сравнения каждой символьной последовательности из одного множества со всеми строками из другого. В работе представлен метод, решающий данную задачу существенно быстрее, чем прямой перебор сравнений со всеми строками подряд. Для ускорения процедуры предлагается оригинальный алгоритм, объединяющий использование префиксного дерева и стандартного алгоритма динамического программирования для поиска редакционного расстояния (метрики Левенштейна) между строками.

- [1] Яковлев П. А. Метод быстрого множественного попарного выравнивания на основе префиксных деревьев // Доклады Академии наук, Москва: РАН, 2018 (в печати).

Fast Trie-based method for multiple pairwise sequence alignment

*Yakovlev Pavel*¹★

yakovlev@biocad.ru

¹Saint-Petersburg, BIOCAD

For more than twenty years monoclonal antibodies are the main origin of novel biopharmaceuticals for oncology and autoimmune diseases. The high efficiency of such type of therapy is caused by exceptional properties of these molecules, such as high variability that allows to obtain molecules that bind to every treatment target. Despite this variability antibodies have a very conservative organization of amino acid sequence. Unceasing interest in the use of antibodies leads to the use of advanced biological methods of analysis, including next generation sequencing (NGS). So millions of sequences of different antibodies are obtained from every NGS experiment.

Sequencing data analysis includes antibody annotation – classification of sequences to one of tens thousands classes and functional regions determination. This procedure requires a comparison for each sequence from some NGS-obtained set with every single sequence from some other one. In this work I present a method that makes possible to solve this task sufficiently faster than direct use of pairwise comparisons. The prefix tree (or trie) is combined with standard dynamic programming algorithms to find edit distances for each pair of sequences from two different sets.

- [1] *Yakovlev P.* Fast trie-based method for multiple pairwise sequence alignment // Doklady Mathematics, Moscow: Pleiades Publishing, 2018 (in press).

Вычислительные технологии решения квазисепарабельных задач безусловной минимизации

*Андреанов Александр Николаевич*¹

and@a5.kiam.ru

*Горнов Александр Юрьевич*²

gornov@icc.ru

Аникин Антон Сергеевич^{2*}

htower@icc.ru

¹Москва, Институт Прикладной Математики им. М.В. Келдыша РАН

²Иркутск, Институт динамики систем и теории управления имени

В.М. Матросова СО РАН

DOI: 10.30826/IDP201816

Термин «квазисепарабельная функция» к настоящему времени не имеет общепринятой однозначной трактовки. В докладе будем называть квазисепарабельной функцией дважды непрерывно дифференцируемую функцию, гессиан которой близок в каком-то смысле к диагональному. Основной идеей предлагаемого подхода является учет специфических особенностей задач безусловной оптимизации, формулируемых в отношении квазисепарабельных функций.

Среди множества существующих алгоритмов оптимизации, особое место заняли метод Б.Т. Поляка (1969 г.), а также немонотонный метод Varzilai-Vorwein (1988 г.). Главными особенностями этих методов является простая структура отдельной итерации и отсутствие вычислительно затратной процедуры одномерного поиска. Предложенная авторами модификация метода Б.Т. Поляка, сохраняющая монотонную сходимость на итерациях, показала в итоге наилучшие результаты на исследуемых задачах больших размерностей.

Программная реализация алгоритмов выполнена на языке C++ с применением технологии параллельных вычислений MPI. Вычислительные эксперименты проведены на высокопроизводительной вычислительной системе К-100 ИПМ РАН. Представлены результаты решения модельных задач оптимизации квазисепарабельных функций с размерностями до 10^{11} переменных.

Работа поддержана грантом РФФИ № 16-07-00664.

- [1] *Andrianov A.N., Anikin A.S., Bychkov I.V., Gornov A.Yu.* Numerical Solution of Huge-Scale Quasiseparable Optimization Problems // Lobachevskii Journal of Mathematics. 2017. Vol. 38. No. 5. P. 870–873.

Computational technologies for solving quasi-separable problems of unconditional minimization

*Andrianov Alexander*¹

and@a5.kiam.ru

*Gornov Alexander*²

gornov@icc.ru

Anikin Anton^{2*}

htower@icc.ru

¹Moscow, Keldysh Institute of Applied Mathematics RAS

²Irkutsk, Matrosov Institute for System Dynamics and Control Theory
SB RAS

The term ‘‘quasi-separable function’’ by the present time does not have a generally accepted interpretation. In this paper we shall call a quasiseparable function a twice continuously differentiable function, the Hessian of which is close to the diagonal. The main idea of the proposed approach is to take into account the specific features of the unconditional optimization problems formulated with respect to quasi-separable functions.

Among the many existing optimization algorithms a special place was taken by the method of B.T. Polyak (1969), and also the non-monotonic method of Barzilai-Borwein (1988). The main features of these methods is the simple structure of a single iteration and the absence of a computationally expensive one-dimensional search procedure. The modification proposed by the authors of Polyak’s method, which preserves monotonous convergence on iterations, showed the best results for the large dimensions problems.

Software implementation of algorithms was realized in C++ language using parallel MPI technology. Computational experiments were carried out on a high-performance computer system K-100 IAM RAS. It is presented the results of solving model optimization problems for quasi-separable functions with dimensions up to 10^{11} variables.

This research is funded by RFBR, grant 16-07-00664.

- [1] *Andrianov A.N., Anikin A.S., Bychkov I.V., Gornov A.Yu.* Numerical Solution of Huge-Scale Quasiseparable Optimization Problems // Lobachevskii Journal of Mathematics. 2017. Vol. 38. No. 5. P. 870–873.

О некоторых полиномиально разрешимых подклассах сетевой задачи размещения

Гимади Эдуард Хайрутдинович^{1,2} gimadi@math.nsc.ru
Курочкина Анна Анатольевна^{3*} a.potapova@gmail.com
*Нагорная Елена Александровна*² nitem91@ya.ru
*Шамардин Юрий Владиславович*¹ orlab@math.nsc.ru
*Шевяков Александр Сергеевич*² shevash.97@gmail.com

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

²Новосибирский государственный университет, Новосибирск

³СибГУТИ, Новосибирск

DOI: 10.30826/IDP201817

Задача размещения (Facility Location Problem, FLP) — одна из типовых в исследовании операций. В общем случае она NP-трудна как в многоэтапной, так и в одноэтапной постановках. В случае древовидных сетей задача решается за время $O(mn)$, где n — число пунктов спроса, m — число возможных мест размещения предприятий. Однако, при ограничениях на объемы производства предприятий задача труднорешаема даже на линейной сети. В данном докладе рассмотрена задача Uniform FLP на путевом графе в предположении, что объемы производства предприятий одинаковы. Трудоемкость $O(m^4n^2)$ известного ранее алгоритма улучшена в m раз [1]. Для NP-трудной в общем случае многоэтапной задачи в работе [2] показано, что результат Трубина и Шарифова (1992) о полиномиальной разрешимости многоэтапной задачи на древовидной сети не корректен. Наряду с этим в [2] приведен ряд примеров полиномиально разрешимых случаев задачи на древовидных сетях. Таким образом, вопрос об NP-трудности задачи на древовидной сети в случае числа этапов больше двух остается открытым. Работа частично поддержана грантом РФФИ № 16-07-00168 и грантом РАН № 0314-2016-0015.

- [1] Gimadi E.Kh., Kurochkina A.A. Time Complexity of the Known Algorithm to Solve the Uniform Hard Capacities Facility Location Problem on Path Graphs // Lecture Notes in Computer Science, Springer, 2018 (accepted).
- [2] Edward Gimadi and Yury Shamardin. On Multi-level Network Facility Location Problem // CEUR-WS.org, online <http://ceur-ws.org>, 2018 (в печати)

On some polynomially solvable subclasses of Network FLP

Gimadi Edward^{1,2}

`gimadi@math.nsc.ru`

*Kurochkina Anna*³★

`a.potapova@gmail.com`

*Nagornaya Elena*²

`nitem91@ya.ru`

*Shamardin Yury*¹

`orlab@math.nsc.ru`

*Shevyakov Alexandr*²

`shevash.97@gmail.com`

¹Sobolev Institute of Mathematics, Novosibirsk, Russia

²Novosibirsk State University, Novosibirsk, Russia

³SibGUTI, Novosibirsk, Russia

The Facility Location Problem (FLP) is one of the typical operations research problems in the study. Its The Facility Location Problem (FLP) is one of the typical operations research problems in the study. Its many different modifications, both in a multi-stage and in a one-step setting, are generally NP-hard.

In the case of tree networks the problem is solved in time $O(mn)$, where n is the number of points of demand, m is the number of possible locations for facilities. However, in the presence of restrictions on the facility capacities, the problem is difficult to solve even on a linear network.

In this report, the problem of Uniform FLP on the line graph is considered, assuming that the capacities of all facilities are the same. The complexity $O(m^4n^2)$ of the previously known algorithm we improve in m times [1].

This work was partially supported by RFBR, grant N 16-07-00168 and by the Russian Academy of Science, grant 0314-2016-0015.

- [1] Gimadi E.Kh., Kurochkina A.A. Time Complexity of the Known Algorithm to Solve the Uniform Hard Capacities Facility Location Problem on Path Graphs // Lecture Notes in Computer Science, Springer, 2018 (accepted).
- [2] Edward Gimadi and Yury Shamardin. On Multi-level Network Facility Location Problem // CEUR-WS.org, online <http://ceur-ws.org>, 2018 (in press).

О сетевой задаче размещения с ограничениями на пропускные способности и объемы производства

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

*Курочкина Анна Александровна*³ a.potapova@ngs.ru

Цидулко Оксана Юрьевна^{1,2*} tsidulko@math.nsc.ru

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

²Новосибирск, Механико-математический факультет НГУ

³Новосибирск, СибГУТИ

DOI: 10.30826/IDP201818

В классической NP-трудной задаче размещения на сети (Facility location problem или FLP) задан n -вершинный граф $G = (V, E)$, в каждой вершине $i \in V$ находится клиент с объемом спроса $b(i)$ и возможно открытие предприятия стоимостью $f(i)$. Для каждого ребра $e \in E$ известна стоимость $c(e)$ транспортировки единицы товара вдоль e . Требуется разместить предприятия, так чтобы удовлетворить спрос и минимизировать суммарные издержки, связанные с транспортировкой товаров и открытием предприятий. В докладе рассматриваются естественные модификации задачи FLP, в которых объем производства на предприятии, открытом в i -ой вершине, ограничен заданной величиной $a(i)$ (Capacitated FLP или CFLP), или введены ограничения на пропускную способность ребер графа (Restricted FLP или RFLP), или введены ограничения обоих типов (Restricted Capacitated FLP или RCFLP). Предлагается точный псевдополиномиальный алгоритм решения задачи RCFLP для случая, когда исходный граф – дерево, а каждый клиент может быть одновременно обслужен несколькими предприятиями. Для задачи RFLP на линейном графе, как в случае, когда каждый клиент должен быть обслужен ровно одним предприятием, так и в случае, когда допускается обслуживание одного клиента несколькими предприятиями, построены точные алгоритмы с временем работы $O(n^3)$. Работа частично поддержана грантом РФФИ № 18-31-00470.

- [1] *Gimadi E. Kh., Kurochkina A., Tsidulko O.* On Exact Solvability of the Restricted Capacitated Facility Location Problem // CEUR-WS Vol 1987, 2017. — P. 209–216. <http://ceur-ws.org/Vol-1987/paper31.pdf>.

On the Network Restricted Capacitated Facility Location Problem

Gimadi Edward^{1,2}

`gimadi@math.nsc.ru`

*Kurochkina Anna*³

`a.potapova@ngs.ru`

Tsidulko Oaxana^{1,2*}

`tsidulko@math.nsc.ru`

¹Novosibirsk, Sobolev Institute of Mathematics

²Novosibirsk, Department of Mechanics and Mathematics, Novosibirsk State University

³Novosibirsk, Siberian State University of Telecommunications and Information Science

The well-known NP-hard network Facility Location Problem (FLP) states as follows. Given an n -vertex network graph $G = (V, E)$, in each vertex $i \in V$ there is a client with demand $b(i)$ and the cost of opening a facility in i is $f(i)$. For each edge $e \in E$ given cost $c(e)$ of transportation of a unit of product along e . The problem is to open a subset of facilities such that all the clients are served, and the total transportation and facility opening costs are minimized. We consider the natural modifications of this problem, where for each $i \in V$ the facility capacities $a(i)$ are given (Capacitated FLP or CFLP), or the edge capacities of the network are given (Restricted FLP or RFLP), or we have both of these capacity constraints (Restricted Capacitated FLP or RCFLP).

We propose an exact pseudopolynomial algorithm for the tree-network RCFLP, where each client can be served by multiple facilities. For the line-network RFLP, for both cases when each client should be served by only one facility and when each client can be served by multiple facilities, we present exact algorithms with $O(n^3)$ running-time.

The authors were partially supported by RFBR, grant 18-31-00470.

- [1] *Gimadi E. Kh., Kurochkina A., Tsidulko O.* On Exact Solvability of the Restricted Capacitated Facility Location Problem // CEUR-WS Vol 1987, 2017. — P. 209–216. <http://ceur-ws.org/Vol-1987/paper31.pdf>.

О задаче Random MST с верхним ограничением на диаметр остова

Гимади Эдуард Хайрутдинович^{1,2}

gimadi@math.nsc.ru

Истомин Алексей Михайлович^{1*}

alexseyistomin@gmail.com

Шин Екатерина Юрьевна¹

katherine15963@gmail.com

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

²Новосибирский государственный университет, Новосибирск

DOI: 10.30826/IDP201819

Пусть задан полный неориентированный граф $G_n = (V, E)$ с множеством вершин $V = \{1, \dots, n\}$, в котором каждому ребру $e = (i, j) \in E$ приписана стоимость (вес) $w_{ij} \geq 0$, $1 \leq i < j \leq n$. Задача заключается в нахождении в графе G_n остовного дерева D_n минимального веса при условии ограниченности его диаметра. Под диаметром искомого подграфа $G' \in G$ понимается длина максимальной, относительно числа ребер, простой цепи в G' . В общем случае данная задача является NP -трудной.

Ранее эта задача рассматривалась при ограничении на диаметр снизу. В настоящем сообщении предлагается приближенный алгоритм с учетом верхнего ограничения на диаметр искомого остовного дерева. Алгоритм решения задачи состоит из двух этапов. 1. Строим цепь $C(\hat{d}) = (i_1, i_2, \dots, i_{\hat{d}+1})$ из \hat{d} ребер. Вначале в качестве i_1 берем произвольную вершину графа и полагаем $C(0) = (i_1)$. Когда построена цепь $C(k)$, $k \leq \hat{d}$, в качестве i_{k+1} берется вершина вне цепи $C(k)$, ближайшая к i_k . 2. Каждую вершину вне цепи $C(\hat{d})$, соединяем кратчайшим ребром с вершиной, лежащей внутри цепи $C(\hat{d})$. В итоге построено остовное дерево с диаметром, не превышающим \hat{d} .

В докладе предполагается доложить результаты вероятностного анализа работы алгоритма квадратичной трудоемкости на случайных входах, рассматриваемых в предвдущих работах с нижним ограничением на диаметр.

Работа частично поддержана грантом РФФИ № 16-07-00168 и грантом РАН № 0314-2016-0015.

- [1] Edward Gimadi, Alexsey Iistomin and Ekaterina Shin. On Bounded Diameter MST Problem on Random Instances // CEUR-WS.org, online <http://ceur-ws.org>, 2018 (accepted)

On Bounded Diameter MST Problem on Random Instances

Gimadi Edward^{1,2}

`gimadi@math.nsc.ru`

Istomin Alexey^{1*}

`alexeyistomin@gmail.com`

*Shin Ekaterina*¹

`katherine15963@gmail.com`

¹Sobolev Institute of Mathematics, Novosibirsk, Russia

²Novosibirsk State University, Novosibirsk, Russia

We give an approximation deterministic algorithm for solving the Random bounded diameter minimum spanning tree (BDMST) problem on an undirected graph. The algorithm has a quadratic time complexity. A probabilistic analysis was performed under conditions that edge weights of given graph are identically independent uniformly distributed random variables on an interval $(a_n; b_n)$. Conditions of asymptotic optimality are presented.

This work was partially supported by RFBR, grant N 16-07-00168 and by the Russian Academy of Science, grant 0314-2016-0015.

- [1] Edward Gimadi, Alexey Istomin and Ekaterina Shin. On Bounded Diameter MST Problem on Random Instances // CEUR-WS.org, online <http://ceur-ws.org>, 2018 (accepted)

О реализации точного метода решения задачи объемно-календарного планирования сетевого проекта с ограниченными ресурсами

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

*Гончаров Евгений Николаевич*¹ gon@math.nsc.ru

Мишин Дмитрий Валентинович^{1,2}★ mishindv@gmail.com

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

²Новосибирск, Новосибирский государственный университет

DOI: 10.30826/IDP201820

Рассматривается задача объемно-календарного планирования в условиях ограниченных ресурсов с критерием минимизации срока окончания проекта. Прерывания работ не допускаются

Мы предлагаем два метода решения задачи с возобновимыми ресурсами. Первым из них является точный метод ветвей и границ с новой схемой ветвления на основе представления сетевого графика в виде списка работ. Эффективное построение нижней границы основано на полиномиальной разрешимости задачи в случае замены возобновимого типа ресурса на кумулятивный (складируемый).

Во втором методе решения задачи с возобновимыми ресурсами мы используем метаэвристику.

Мы представляем результаты численных экспериментов, иллюстрирующих качество предлагаемого алгоритма. Для проведения расчетов были использованы тестовых экземпляров из известной библиотеки PSPLIB. Численные эксперименты продемонстрировали конкурентоспособность алгоритма. Для нескольких экземпляры из набора данных j120M были найдены рекордно лучшие решения по сравнению с ранее полученными [1].

Работа частично поддержана грантами РФФИ № 16-07-00168, № 16-07-00829 и грантом РАН № 0314-2016-0015.

- [1] *E.Kh. Gimadi, E.N. Goncharov, D.V. Mishin* On some implementations of solving the resource constrained project scheduling problems // Yugoslav Journal of Operations Research, Beograd, 2018 (в печати).

<http://yujor.fon.bg.ac.rs/index.php/yujor/article>

On Some Realizations of Solving the Resource Constrained Project Scheduling Problems

Gimadi Edward^{1,2}

`gimadi@math.nsc.ru`

*Goncharov Eugeny*¹

`gon@math.nsc.ru`

*Mishin Dmitry*¹★

`mishindmv@gmail.com`

¹Novosibirsk, Sobolev Institute of Mathematics

²Novosibirsk, Novosibirsk State University

We consider the resource-constrained project scheduling problem with respect to the makespan minimization criterion. The problem accounts for technological constraints of activities precedence together with resource constraints. Activities preemptions are not allowed. We consider renewable and cumulative resources.

We introduce that the problem with cumulative resource constraints can be solved in strongly polynomial time. We propose two methods for solving the problem with renewable resources. One of them is the exact branch and bound algorithm with a new branching scheme based on the presentation of a schedule in the form of a activity list.

We use two variants of constructing the lower bound. In another method we use metaheuristics.

We present results of numerical experiments illustrating quality of proposed algorithm. The test instances were used from the library of test instances PSPLIB. Numerical experiments demonstrated algorithm's competitiveness. We have found the best solutions for a few instances from the dataset j120 [1].

This research is partially supported by RFBR, grants 16-07-00829, 16-07-00168 and by the Russian Academy of Science, grant 0314-2016-0015.

- [1] *E.Kh. Gimadi, E.N. Goncharov, D.V. Mishin* On some implementations of solving the resource constrained project scheduling problems // Yugoslav Journal of Operations Research, Beograd, 2018 (in press). <http://yujor.fon.bg.ac.rs/index.php/yujor/article>

Алгоритмы облачной аппроксимации невыпуклых множеств в конечномерном пространстве

Горнов Александр Юрьевич

gornov@icc.ru

Иркутск, Институт динамики систем и теории управления имени

В.М. Матросова СО РАН

DOI: 10.30826/IDP201821

Одной из популярных в приложениях задач является задача аппроксимации невыпуклых множеств в конечномерном пространстве. Проблемы такого типа возникают в целом ряде областей прикладной математики — в задачах фазового оценивания, задачах невыпуклой оптимизации, задачах поиска глобального экстремума функционалов, задачах синтеза оптимального управления и многих других. Однако работ, посвященных данной проблеме, опубликовано удивительно мало. В докладе обсуждаются набор алгоритмов аппроксимации невыпуклого множества, основанных на идее «облачных аппроксимаций» — приближения сетками квазислучайного характера («нерегулярными сетками»). Рассматриваемые алгоритмы ориентированы на получение информации о геометрии множества в пространствах, вообще говоря, произвольной размерности. С применением предложенного подхода реализованы методы для теоретико-множественных операций объединения, пересечения, дополнения и выпукления, аппроксимации границ множества. Для специального случая рассматриваемой проблемы — при наличии потенциальной функции, определенной на искомом множестве — реализованы также алгоритмы кластеризации для оценки областей с «низкими потенциалом», алгоритмы оценки плотности облачной аппроксимации в выявленных кластерах и другие. В качестве приложения обсуждается применение предложенных алгоритмов для исследования свойств множества достижимости управляемой нелинейной динамической системы и оценки «областей залегания» глобального экстремума многомерной невыпуклой функции. Работа поддержана грантом РФФИ №18-07-00587.

- [1] *Gornov A., Zarodnyuk T., Anikin A., Finkelstein E.* Technology of nonlocal search in optimal control problems, based on the property of hidden convexity // *Journal of Global Optimization*, Springer, 2018 (в печати).

Algorithms of cloud approximation for nonconvex sets in a finite-dimensional space

Gornov Alexander

gornov@icc.ru

Irkutsk, Matrosov Institute for System Dynamics and Control Theory of SB RAS

One of the popular problems in applications is the problem of approximating nonconvex sets in a finite-dimensional space. Problems of this type arise in a number of areas of applied mathematics — in problems of phase estimation, nonconvex optimization, finding a global extremum of functionals, optimal control synthesis, and many others. However, few papers devoted to this problem have been published.

The paper discusses a set of algorithms for approximating a nonconvex set based on the idea of “cloud approximations” — approximation by grids of a quasi-random nature (“irregular grids”). The algorithms under consideration are oriented to obtaining information about the geometry of a set in spaces, generally speaking, of arbitrary dimension. With the application of the proposed approach, methods are implemented for set-theoretic operations of union, intersection, complementing and convexification, approximation of the boundaries of a set. For the special case of the problem under consideration, in the presence of a potential function defined on the desired set, clustering algorithms for estimating regions with “low potential”, algorithms for estimating the density of cloud approximation in the clusters detected, algorithms for estimating the distances between the cluster centers, and others are realized.

The use of the proposed algorithms for investigating the properties of the reachable set of a controllable nonlinear dynamical system and estimating the “location areas” of a global extremum of a multidimensional nonconvex function are discussed as an application.

This research is funded by RFBR, grant 18-07-00587.

- [1] *Gornov A. Zarodnyuk T. Anikin A., Finkelstein E.* Technology of non-local search in optimal control problems, based on the property of hidden convexity // *Journal of Global Optimization*, Springer, 2018 (in press).

Точные алгоритмы для специальных случаев двух труднорешаемых задач 2-кластеризации

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

Хандеев Владимир Ильич^{1,2} khandeev@math.nsc.ru

Панасенко Анна Владимировна^{1,2} a.v.panasenko@math.nsc.ru

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

²Новосибирск, Новосибирский государственный университет

DOI: 10.30826/IDP201822

В работе рассматриваются две оптимизационные задачи.

Задача 1. Дано: N -элементное множество \mathcal{Y} точек в евклидовом пространстве размерности d и число $\alpha \in (0, 1)$. Найти: подмножество $\mathcal{C} \subset \mathcal{Y}$ наибольшей мощности такое, что

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

где $\bar{y}(\mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum_{y \in \mathcal{C}} y$ и $\bar{y}(\mathcal{Y}) = \frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}} y$ — центроиды \mathcal{C} и \mathcal{Y} .

Задача 2. Дано: N -элементное множество \mathcal{Y} точек в евклидовом пространстве размерности d . Найти: подмножество $\mathcal{C} \subset \mathcal{Y}$ наибольшей мощности такое, что

$$|\mathcal{C}| \sum_{y \in \mathcal{C}} \|y - \bar{y}(\mathcal{C})\|^2 + |\mathcal{Y} \setminus \mathcal{C}| \sum_{y \in \mathcal{Y} \setminus \mathcal{C}} \|y\|^2 \leq \alpha N \sum_{y \in \mathcal{Y}} \|y - \bar{y}(\mathcal{Y})\|^2.$$

Обе задачи индуцируются проблемами редактирования данных. В работе установлено, что обе задачи NP-трудны в сильном смысле. Обоснованы точные алгоритмы для случаев этих задач, в которых точки входного множества имеют целочисленные координаты. Если размерность пространства фиксирована, то оба предложенных алгоритма псевдополиномиальны.

Исследование Задачи 1 поддержано грантом РФФ 16-11-10041. Исследование Задачи 2 поддержано грантами РФФИ 16-07-00168, 18-31-00398-мол-а, а также грантом РАН (проект 0314-2016-0015).

[1] *Kelmanov A., Khandeev V., Panasenko A.* Exact Algorithms for the Special Cases of Two Hard to Solve Problems of Searching for the Largest Subset // Lecture Notes in Computer Science, Springer, 2018 (в печати).

<http://aistconf.org/schedule/>

Exact Algorithms for the Special Cases of Two Hard to Solve Problems

Kelmanov Alexander^{1,2*}

kelm@math.nsc.ru

Khandeev Vladimir^{1,2}

khandeev@math.nsc.ru

Panasenko Anna^{1,2}

a.v.panasenko@math.nsc.ru

¹Novosibirsk, Sobolev Institute of Mathematics

²Novosibirsk, Novosibirsk State University

We consider two optimization problems.

Problem 1. *Given an N -element set \mathcal{Y} of points in d -dimensional Euclidean space and a number $\alpha \in (0, 1)$. Find a subset $\mathcal{C} \subset \mathcal{Y}$ of the largest cardinality such that*

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

where $\bar{y}(\mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum_{y \in \mathcal{C}} y$ and $\bar{y}(\mathcal{Y}) = \frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}} y$ are the centroids of the subset \mathcal{C} and the given set \mathcal{Y} , respectively.

Problem 2. *Given an N -element set \mathcal{Y} of points in d -dimensional Euclidean space and a number $\alpha \in (0, 1)$. Find a subset $\mathcal{C} \subset \mathcal{Y}$ of the largest cardinality such that*

$$|\mathcal{C}| \sum_{y \in \mathcal{C}} \|y - \bar{y}(\mathcal{C})\|^2 + |\mathcal{Y} \setminus \mathcal{C}| \sum_{y \in \mathcal{Y} \setminus \mathcal{C}} \|y\|^2 \leq \alpha N \sum_{y \in \mathcal{Y}} \|y - \bar{y}(\mathcal{Y})\|^2.$$

Both problems model the data editing problems. In this paper, we prove the strong NP-hardness of these problems and present exact algorithms for their special cases (integer-valued coordinates of the input points); if the space dimension is fixed, both algorithms are pseudopolynomial.

The study of Problem 1 was supported by the RSF, project 16-11-10041. The study of Problem 2 was supported by the RFBR, projects 16-07-00168, 18-31-00398, and by the Russian Academy of Science, project 0314-2016-0015.

[1] *Kelmanov A., Khandeev V., Panasenko A.* Exact Algorithms for the Special Cases of Two Hard to Solve Problems of Searching for the Largest Subset // Lecture Notes in Computer Science, Springer, 2018 (in press).

<http://aistconf.org/schedule/>

Рандомизированные алгоритмы для некоторых задач кластеризации

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

Хандеев Владимир Ильич^{1,2} khandeev@math.nsc.ru

Панасенко Анна Владимировна^{1,2*} a.v.panasenko@math.nsc.ru

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

²Новосибирск, Новосибирский государственный университет

DOI: 10.30826/IDP201823

Рассматриваются NP-трудные в сильном смысле задачи:

Задача 1. Дано: N -элементное множество \mathcal{Y} точек из \mathbb{R}^q и натуральное число M . Найти: подмножество $\mathcal{C} \subset \mathcal{Y}$ мощности M такое, что

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

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

Задача 2. Дано: N -элементное множество \mathcal{Y} точек из \mathbb{R}^q и натуральное число M . Найти: разбиение множества \mathcal{Y} на два подмножества \mathcal{C} (мощности M) и $\mathcal{Y} \setminus \mathcal{C}$ (мощности $N - M$) такое, что

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

Обе задачи индуцируются проблемами кластеризации данных. В работе предложены рандомизированные алгоритмы, которые находят приближенные решения задач за время, линейно зависящее от размерности пространства и от мощности входного множества. Найдены условия, при которых алгоритмы полиномиальны и асимптотически точны.

Исследование Задачи 1 поддержано грантом РФФ 16-11-10041. Исследование Задачи 2 поддержано грантами РФФИ 16-07-00168, 18-31-00398-мол-а, а также грантом РАН (проект 0314-2016-0015).

- [1] *Kelmanov A., Khandeev V., Panasenko A.* Randomized Algorithms for Some Clustering Problems // Communications in Computer and Information Science. Vol. 871, P. 109–119, 2018.

Randomized Algorithms for Some Clustering Problems

Kelmanov Alexander^{1,2}

kelm@math.nsc.ru

Khandeev Vladimir^{1,2}

khandeev@math.nsc.ru

Panasenko Anna^{1,2}*

a.v.panasenko@math.nsc.ru

¹Novosibirsk, Sobolev Institute of Mathematics

²Novosibirsk, Novosibirsk State University

We consider the following strongly NP-hard problems.

Problem 1. *Given an N -element set \mathcal{Y} of points in \mathbb{R}^q , and a positive integer M . Find a subset $\mathcal{C} \subseteq \mathcal{Y}$ of size M minimizing the value of*

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

where $\bar{y}(\mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum_{y \in \mathcal{C}} y$ is the centroid of \mathcal{C} .

Problem 2. *Given an N -element set \mathcal{Y} of points in \mathbb{R}^q , and positive integer number M . Find a partition of \mathcal{Y} into two non-empty clusters \mathcal{C} (of size M) and $\mathcal{Y} \setminus \mathcal{C}$ (of size $N - M$), such that*

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

Both problems induced by the data clustering. In this paper, we present randomized algorithms for these problems. Our algorithms allow one to find an approximate solution in a time that is linear in the space dimension and the input size of the problems for given upper bounds of the relative error, failure probability and for an established parameter value. The conditions are found under which the algorithms are polynomial and asymptotically exact.

The study of Problem 1 was supported by the RSF, project 16-11-10041. The study of Problem 2 was supported by the RFBR, projects 16-07-00168, 18-31-00398, and by the Russian Academy of Science, project 0314-2016-0015.

- [1] *Kelmanov A., Khandeev V., Panassenko A.* Randomized Algorithms for Some Clustering Problems // Communications in Computer and Information Science. Vol. 871, P. 109–119, 2018.

Точные алгоритмы для двух задач поиска наибольшего подмножества и наибольшей подпоследовательности

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

Хамидуллин Сергей Асгадулович^{1*} kham@math.nsc.ru

Хандеев Владимир Ильич^{1,2} khandeev@math.nsc.ru

Пяткин Артем Валерьевич^{1,2} artem@math.nsc.ru

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

²Новосибирск, Новосибирский государственный университет

DOI: 10.30826/IDP201824

В работе предложены точные алгоритмы для целочисленных вариантов следующих NP-трудных в сильном смысле задач.

Problem 1. Дано: множество $\mathcal{Y} = \{y_1, \dots, y_N\}$ точек из \mathbb{R}^q и число $\alpha \in (0, 1)$. Найдти подмножество $\mathcal{C} \subset \mathcal{Y}$ наибольшей мощности такое, что $\sum_{y \in \mathcal{C}} \|y - \bar{y}(\mathcal{C})\|^2 \leq \alpha \sum_{y \in \mathcal{Y}} \|y - \bar{y}(\mathcal{Y})\|^2$, где $\bar{y}(\mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum_{y \in \mathcal{C}} y$ и $\bar{y}(\mathcal{Y}) = \frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}} y$ — центроиды \mathcal{C} и \mathcal{Y} .

Problem 2. Дано: последовательность $\mathcal{Y} = (y_1, \dots, y_N)$ точек из \mathbb{R}^q , натуральные числа T_{\min} , T_{\max} , и число $\alpha \in (0, 1)$. Найдти: подмножество $\mathcal{M} = \{n_1, \dots, n_M\} \subseteq \mathcal{N} = \{1, \dots, N\}$ наибольшей мощности такое, что $T_{\min} \leq n_m - n_{m-1} \leq T_{\max} \leq N$, $m = 2, \dots, M$, и

$$\sum_{j \in \mathcal{M}} \|y_j - \bar{y}(\mathcal{M})\|^2 \leq \alpha \sum_{j \in \mathcal{N}} \|y_j - \bar{y}(\mathcal{N})\|^2,$$

где $\bar{y}(\mathcal{M}) = \frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} y_i$ и $\bar{y}(\mathcal{N}) = \frac{1}{N} \sum_{i \in \mathcal{N}} y_i$ — центроиды $\{y_i \in \mathcal{Y} \mid i \in \mathcal{M}\}$ и $\{y_i \in \mathcal{Y} \mid i \in \mathcal{N}\}$.

Обе задачи моделируют проблемы редактирования данных. Предложенные алгоритмы псевдополиномиальны, если размерность пространства ограничена сверху константой.

Работа частично поддержана грантами РФФИ 16-07-00168, 18-31-00398 и грантом РАН 0314-2016-0015.

- [1] Kelmanov A., Khamidullin S., Khandeev V., Pyatkin A. Exact Algorithms for Two Quadratic Euclidean Problems of Searching for the Largest Subset and Longest Subsequence // Lecture Notes in Computer Science, Springer, 2018 (в печати).

Exact Algorithms for Two Problems of Searching for the Largest Subset and Longest Subsequence

Kelmanov Alexander^{1,2}

kelm@math.nsc.ru

Khamidullin Sergey^{1*}

kham@math.nsc.ru

Khandeev Vladimir^{1,2}

khandeev@math.nsc.ru

Pyatkin Artem^{1,2}

artem@math.nsc.ru

¹Novosibirsk, Sobolev Institute of Mathematics

²Novosibirsk, Novosibirsk State University

In this paper we present exact algorithms for the special cases (the instances are integer) of the following strongly NP-hard problems.

Problem 1. *Given a set $\mathcal{Y} = \{y_1, \dots, y_N\}$ of points in \mathbb{R}^q , and a number $\alpha \in (0, 1)$. Find a subset $\mathcal{C} \subset \mathcal{Y}$ of largest cardinality such that*

$$\sum_{y \in \mathcal{C}} \|y - \bar{y}(\mathcal{C})\|^2 \leq \alpha \sum_{y \in \mathcal{Y}} \|y - \bar{y}(\mathcal{Y})\|^2,$$

where $\bar{y}(\mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum_{y \in \mathcal{C}} y$ and $\bar{y}(\mathcal{Y}) = \frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}} y$ are the centroids of \mathcal{C} and \mathcal{Y} , respectively.

Problem 2. *Given a sequence $\mathcal{Y} = (y_1, \dots, y_N)$ of points in \mathbb{R}^q , some positive integer numbers T_{\min} , T_{\max} , and a number $\alpha \in (0, 1)$. Find a subset $\mathcal{M} = \{n_1, \dots, n_M\} \subseteq \mathcal{N} = \{1, \dots, N\}$ of largest size such that $T_{\min} \leq n_m - n_{m-1} \leq T_{\max} \leq N$, $m = 2, \dots, M$, and*

$$\sum_{j \in \mathcal{M}} \|y_j - \bar{y}(\mathcal{M})\|^2 \leq \alpha \sum_{j \in \mathcal{N}} \|y_j - \bar{y}(\mathcal{N})\|^2,$$

where $\bar{y}(\mathcal{M}) = \frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} y_i$ and $\bar{y}(\mathcal{N}) = \frac{1}{N} \sum_{i \in \mathcal{N}} y_i$ are the centroids of $\{y_i \in \mathcal{Y} \mid i \in \mathcal{M}\}$ and $\{y_i \in \mathcal{Y} \mid i \in \mathcal{N}\}$ respectively.

Both problems can be treated as data editing problems. If the space dimension is bounded by some constant, our algorithms run in a pseudopolynomial time.

This research is partially funded by RFBR, grants 16-07-00168, 18-31-00398, and by the Russian Academy of Science, grant 0314-2016-0015.

- [1] *Kelmanov A., Khamidullin S., Khandeev V., Pyatkin A.* Exact Algorithms for Two Quadratic Euclidean Problems of Searching for the Largest Subset and Longest Subsequence // Lecture Notes in Computer Science, Springer, 2018 (in press).

<http://www.caopt.com/LION12/program.php>.

Точный алгоритм поиска кластера наибольшего размера в задаче 2-кластеризации целочисленной последовательности

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

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

*Хамидуллин Сергей Асгадуллович*¹ kham@math.nsc.ru

Хандеев Владимир Ильич^{1,2} khandeev@math.nsc.ru

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

²Новосибирск, Новосибирский государственный университет

DOI: 10.30826/IDP201825

В работе рассматривается следующая

Задача. Дано: последовательность $\mathcal{Y} = (y_1, \dots, y_N)$ точек в евклидовом пространстве размерности d , натуральные числа T_{\min} , T_{\max} , и число $\alpha \in (0, 1)$. *Найти:* подмножество $\mathcal{M} = \{n_1, \dots, n_M\} \subseteq \mathcal{N} = \{1, \dots, N\}$ наибольшего размера такое, что $T_{\min} \leq n_m - n_{m-1} \leq T_{\max} \leq N$, $m = 2, \dots, M$, и $\sum_{j \in \mathcal{M}} \|y_j - \bar{y}(\mathcal{M})\|^2 + \sum_{j \in \mathcal{N} \setminus \mathcal{M}} \|y_j\|^2 \leq \alpha \sum_{j \in \mathcal{N}} \|y_j - \bar{y}(\mathcal{N})\|^2$, где $\bar{y}(\mathcal{M}) = \frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} y_i$ и $\bar{y}(\mathcal{N}) = \frac{1}{N} \sum_{i \in \mathcal{N}} y_i$ — центры (геометрические центры) $\{y_i \in \mathcal{Y} \mid i \in \mathcal{M}\}$ and $\{y_i \in \mathcal{Y} \mid i \in \mathcal{N}\}$ соответственно.

Задача моделирует, в частности, помехоустойчивый поиск во временной последовательности результатов измерений подмножества максимального размера, состоящего из похожих между собой объектов.

В работе доказана сильная NP-трудность задачи и предложен точный алгоритм для случая задачи, в котором координаты точек целочисленны. Если размерность пространства ограничена сверху константой, то предложенный алгоритм псевдополиномиален.

Работа частично поддержана грантами РФФИ 16-07-00168, 18-31-00398 и грантом РАН 0314-2016-0015.

- [1] *Alexander Kelmanov, Artem Pyatkin, Sergey Khamidullin, Vladimir Khandeev.* An Exact Algorithm of Searching for the Largest Size Cluster in an Integer Sequence 2-Clustering Problem // Communications in Computer and Information Science, 2018 (в печати).
<http://agora.guru.ru/display.php?conf=optima-2018>.

An Exact Algorithm of Searching for the Largest Size Cluster in an Integer Sequence 2-Clustering Problem

Kelmanov Alexander^{1,2}

kelm@math.nsc.ru

Pyatkin Artem^{1,2}*

artem@math.nsc.ru

*Khamidullin Sergey*¹

kham@math.nsc.ru

Khandeev Vladimir^{1,2}

khandeev@math.nsc.ru

¹Novosibirsk, Sobolev Institute of Mathematics

²Novosibirsk, Novosibirsk State University

In this paper we consider following

Problem 1. *Given a sequence $\mathcal{Y} = (y_1, \dots, y_N)$ of points in d -dimensional Euclidean space, positive integers T_{\min}, T_{\max} , and a number $\alpha \in (0, 1)$. Find a subset $\mathcal{M} = \{n_1, \dots, n_M\} \subseteq \mathcal{N} = \{1, \dots, N\}$ of largest size such that $T_{\min} \leq n_m - n_{m-1} \leq T_{\max} \leq N$, $m = 2, \dots, M$, and*

$$\sum_{j \in \mathcal{M}} \|y_j - \bar{y}(\mathcal{M})\|^2 + \sum_{j \in \mathcal{N} \setminus \mathcal{M}} \|y_j\|^2 \leq \alpha \sum_{j \in \mathcal{N}} \|y_j - \bar{y}(\mathcal{N})\|^2,$$

where $\bar{y}(\mathcal{M}) = \frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} y_i$ and $\bar{y}(\mathcal{N}) = \frac{1}{N} \sum_{i \in \mathcal{N}} y_i$ are the centroids of $\{y_i \in \mathcal{Y} \mid i \in \mathcal{M}\}$ and $\{y_i \in \mathcal{Y} \mid i \in \mathcal{N}\}$ respectively.

The problem, in particular, models the noise-prove search for the maximum subset of objects close to each other in the set of time-ordered measurement results.

In this paper, the strong NP-hardness of the problem is shown and an exact algorithm is suggested for the case of integer coordinates of input points. If the space dimension is bounded by some constant this algorithm runs in a pseudopolynomial time.

This research is partially funded by RFBR, grants 16-07-00168, 18-31-00398, and by the Russian Academy of Science, grant 0314-2016-0015.

- [1] *Alexander Kelmanov, Artem Pyatkin, Sergey Khamidullin, Vladimir Khandeev.* An Exact Algorithm of Searching for the Largest Size Cluster in an Integer Sequence 2-Clustering Problem // Communications in Computer and Information Science, 2018 (in press).

<http://agora.guru.ru/display.php?conf=optima-2018>.

Приближенный полиномиальный алгоритм для задачи поиска подпоследовательности наибольшей длины в конечной последовательности точек евклидова пространства

Кельманов Александр Васильевич^{1,2} kelm@math.nsc.ru
Пяткин Артем Валерьевич^{1,2} artem@math.nsc.ru
*Хамидуллин Сергей Асгадулович*¹ kham@math.nsc.ru
Хандеев Владимир Ильич^{1,2*} khandeev@math.nsc.ru
*Шенмайер Владимир Владимирович*¹ shenmaier@mail.ru
*Шамардин Юрий Владиславович*¹ orlab@math.nsc.ru

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

²Новосибирск, Новосибирский государственный университет

DOI: 10.30826/IDP201826

В работе рассматривается следующая экстремальная

Задача. Дано: последовательность $\mathcal{Y} = (y_1, \dots, y_N)$ точек из \mathbb{R}^q , натуральные числа T_{\min} , T_{\max} и число $\alpha \in (0, 1)$. Найти: подмножество $\mathcal{M} = \{n_1, \dots, n_M\} \subseteq \mathcal{N} = \{1, \dots, N\}$ наибольшей мощности такое, что

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

$$\sum_{j \in \mathcal{M}} \|y_j - \bar{y}(\mathcal{M})\|^2 \leq \alpha \sum_{j \in \mathcal{N}} \|y_j - \bar{y}(\mathcal{N})\|^2,$$

где $\bar{y}(\mathcal{M}) = \frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} y_i$ и $\bar{y}(\mathcal{N}) = \frac{1}{N} \sum_{i \in \mathcal{N}} y_i$.

Задача имеет приложения, в частности, в data editing, data cleaning, data mining, machine learning. В настоящей работе установлено, что она NP-трудна в сильном смысле и для ее решения предложен полиномиальный приближенный алгоритм с оценкой точности, близкой к 1/2.

Работа поддержана грантами РФФИ 16-07-00168, 18-31-00398-мол-а, а также грантом РАН (проект 0314-2016-0015).

- [1] *Kelmanov A., Pyatkin A., Khamidullin S., Khandeev V., Shamardin Y.V., Shenmaier V.* An Approximation Polynomial Algorithm for a Problem of Searching for the Longest Subsequence in a Finite Sequence of Points in Euclidean Space. Communications in Computer and Information Science, Vol. 871, P. 120–130, 2018.

An Approximation Polynomial Algorithm for a Problem of Searching for the Longest Subsequence in a Finite Sequence of Points in Euclidean Space

Alexander Kelmanov^{1,2}

kelm@math.nsc.ru

Artem Pyatkin^{1,2}

artem@math.nsc.ru

Sergey Khamidullin¹

kham@math.nsc.ru

Vladimir Khandeev^{1,2}*

khan@math.nsc.ru

Vladimir Shenmaier¹

shenmaier@mail.ru

Yury Shamardin¹

orlab@math.nsc.ru

¹Novosibirsk, Sobolev Institute of Mathematics

²Novosibirsk, Novosibirsk State University

The problem under consideration is stated as follows.

Problem. Given a sequence $\mathcal{Y} = (y_1, \dots, y_N)$ of points in \mathbb{R}^q , some positive integer numbers T_{\min} , T_{\max} , and a number $\alpha \in (0, 1)$. Find a subset $\mathcal{M} = \{n_1, \dots, n_M\} \subseteq \mathcal{N} = \{1, \dots, N\}$ of largest size such that

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

$$\sum_{j \in \mathcal{M}} \|y_j - \bar{y}(\mathcal{M})\|^2 \leq \alpha \sum_{j \in \mathcal{N}} \|y_j - \bar{y}(\mathcal{N})\|^2,$$

where $\bar{y}(\mathcal{M}) = \frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} y_i$ and $\bar{y}(\mathcal{N}) = \frac{1}{N} \sum_{i \in \mathcal{N}} y_i$.

The problem is important for the applications related, in particular, to data editing, data cleaning, data mining, machine learning. In the current paper, we show that the problem is strongly NP-hard and we propose an effective tool for solving it — a polynomial-time approximation algorithm with preciseness bound close to 1/2.

This research was supported by the Russian Foundation for Basic Research, projects 16-07-00168, 18-31-00398, by the Russian Academy of Science, project 0314-2016-0015.

- [1] Kelmanov A., Pyatkin A., Khamidullin S., Khandeev V., Shamardin Y.V., Shenmaier V. An Approximation Polynomial Algorithm for a Problem of Searching for the Longest Subsequence in a Finite Sequence of Points in Euclidean Space. Communications in Computer and Information Science, Vol. 871, P. 120–130, 2018.

NP-трудность некоторых максиминных задач кластеризации

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

Пяткин Артем Валерьевич^{1,2} artem@math.nsc.ru

Хандеев Владимир Ильич^{1,2*} khandeev@math.nsc.ru

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

²Новосибирск, Новосибирский государственный университет

DOI: 10.30826/IDP201827

В работе рассматриваются следующие задачи.

Задача 1. Дано: N -элементное множество \mathcal{Y} точек в d -мерном евклидовом пространстве и число $\alpha \in (0, 1)$. Найти: непустые непересекающиеся подмножества $\mathcal{C}_1, \mathcal{C}_2$ и точки y_1, y_2 во множестве \mathcal{Y} такие, что

$$\min\{|\mathcal{C}_1|, |\mathcal{C}_2|\} \rightarrow \max \quad (1)$$

при ограничениях

$$\sum_{y \in \mathcal{C}_i} \|y - y_i\|^2 \leq \alpha \sum_{y \in \mathcal{Y}} \|y - \bar{y}(\mathcal{Y})\|^2, \quad i = 1, 2,$$

где $\bar{y}(\mathcal{Y}) = \frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}} y$ — центр масс множества \mathcal{Y} .

Задача 2. Дано: N -элементное множество \mathcal{Y} точек в d -мерном евклидовом пространстве, число $\alpha \in (0, 1)$ и точки $z_1, z_2 \in \mathbb{R}^d$. Find непустые непересекающиеся подмножества $\mathcal{C}_1, \mathcal{C}_2$ такие, что имеет место (1) при ограничениях

$$\sum_{y \in \mathcal{C}_i} \|y - z_i\|^2 \leq \alpha \sum_{y \in \mathcal{Y}} \|y - \bar{y}(\mathcal{Y})\|^2, \quad i = 1, 2.$$

Обе задачи моделируют очистку данных от выбросов. В работе установлена NP-трудность Задач 1 и 2, а также их обобщений на случай произвольного числа кластеров.

Работа поддержана грантом РФФИ 16-11-10041.

- [1] Kel'manov A., Khandeev V., Pyatkin A. NP-hardness of Some Max-Min Clusterization Problems // Communications in Computer and Information Science, 2018 (в печати).

<http://agora.guru.ru/display.php?conf=optima-2018>.

NP-hardness of Some Max-Min Clusterization Problems

Alexander Kelmanov^{1,2}

kelm@math.nsc.ru

Artem Pyatkin^{1,2}

artem@math.nsc.ru

Vladimir Khandeev^{1,2}*

khan@math.nsc.ru

¹Novosibirsk, Sobolev Institute of Mathematics

²Novosibirsk, Novosibirsk State University

The problems under consideration are stated as follows.

Problem 1. *Given an N -element set \mathcal{Y} of points in d -dimensional Euclidean space and a number $\alpha \in (0, 1)$. Find non-empty disjoint subsets $\mathcal{C}_1, \mathcal{C}_2$ and points y_1, y_2 in the set \mathcal{Y} such that*

$$\min\{|\mathcal{C}_1|, |\mathcal{C}_2|\} \rightarrow \max \quad (1)$$

under constraints

$$\sum_{y \in \mathcal{C}_i} \|y - y_i\|^2 \leq \alpha \sum_{y \in \mathcal{Y}} \|y - \bar{y}(\mathcal{Y})\|^2, \quad i = 1, 2,$$

where $\bar{y}(\mathcal{Y}) = \frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}} y$ is a centroid of \mathcal{Y} .

Problem 2. *Given an N -element set \mathcal{Y} of points in d -dimensional Euclidean space, a number $\alpha \in (0, 1)$, and points $z_1, z_2 \in \mathbb{R}^d$. Find non-empty disjoint subsets $\mathcal{C}_1, \mathcal{C}_2$ such that (1) holds under constraints*

$$\sum_{y \in \mathcal{C}_i} \|y - z_i\|^2 \leq \alpha \sum_{y \in \mathcal{Y}} \|y - \bar{y}(\mathcal{Y})\|^2, \quad i = 1, 2.$$

Both problems can be treated as data cleaning from outliers. In the current paper, we prove NP-hardness of Problems 1 and 2 and their generalizations for the case of an arbitrary number of clusters.

This research was supported by the RSF, project 16-11-10041.

- [1] *Kelmanov A., Khandeev V., Pyatkin A.* NP-hardness of Some Max-Min Clusterization Problems. Communications in Computer and Information Science, 2018 (in press).

<http://agora.guru.ru/display.php?conf=optima-2018>.

Об одной задаче суммирования элементов, выбираемых из семейства конечных числовых последовательностей

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

Михайлова Людмила Викторовна^{1*} mikh@math.nsc.ru

*Романченко Семен Михайлович*¹ rsm@math.nsc.ru

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

²Новосибирск, Новосибирский государственный университет

DOI: 10.30826/IDP201828

В работе рассматривается следующая

Задача. Дано: семейство $\mathcal{S} = \left\{ \{s_1(n), \dots, s_L(n)\} \mid s_i(n) \in \mathbb{R}_{\geq}, i = 1, \dots, L, n = 1, \dots, N \right\}$ последовательностей и натуральное число J , такие что $JL \leq N$. *Найти:* набор номеров (n_1, \dots, n_{JL}) , где $n_k \in \{1, \dots, N\}$, $k = 1, \dots, JL$, и набор (π_1, \dots, π_J) из J перестановок на L элементах такие, что

$$\sum_{j=1}^J \sum_{i=1}^L s_{\pi_j(i)}(n_{(j-1)L+i}) \rightarrow \min$$

при ограничениях

$$1 \leq n_1 < n_2 < \dots < n_{JL} \leq N$$

на элементы набора (n_1, \dots, n_{JL}) .

Задача индуцируется, в частности, проблемой мониторинга нескольких движущихся объектов при возможных произвольных переупорядочениях (перестановках) этих объектов.

Построен точный полиномиальный алгоритм решения этой задачи, имеющий трудоемкость $\mathcal{O}(N^5)$.

Работа поддержана грантом РФФИ 16-07-00168 и грантом РАН 0314-2016-0015.

- [1] *Kel'manov A., Mikhailova L., and Romanchenko S.* On a Problem of Summing Elements Chosen from a Family of Finite Numerical Sequences // *Lecture Notes in Computer Science*, Springer, 2018 (в печати).

On a Problem of Summing Elements Chosen from a Family of Finite Numerical Sequences

Kelmanov Alexander^{1,2}

kelm@math.nsc.ru

Mikhailova Ludmila^{1*}

mikh@math.nsc.ru

*Romanchenko Semyon*¹

rsm@math.nsc.ru

¹Novosibirsk, Sobolev Institute of Mathematics

²Novosibirsk, Novosibirsk State University

In this paper we consider following

Problem. Given a family $\mathcal{S} = \left\{ \{s_1(n), \dots, s_L(n)\} \mid s_i(n) \in \mathbb{R}_{\geq}, i = 1, \dots, L, n = 1, \dots, N \right\}$ of sequences and a positive integer J such that $JL \leq N$. Find an index tuple (n_1, \dots, n_{JL}) , where $n_k \in \{1, \dots, N\}$, $k = 1, \dots, JL$, and a tuple (π_1, \dots, π_J) of J permutations on L elements such that

$$\sum_{j=1}^J \sum_{i=1}^L s_{\pi_j(i)}(n_{(j-1)L+i}) \rightarrow \min$$

subject to constraints

$$1 \leq n_1 < n_2 < \dots < n_{JL} \leq N.$$

The problem is induced by the remote monitoring of several moving objects with possible arbitrary displacements (permutations) of these objects.

We present an exact polynomial-time algorithm with $\mathcal{O}(N^5)$ running time for this problem.

This research is partially funded by RFBR, grants 16-07-00168 and by the Russian Academy of Science, grant 0314-2016-0015.

- [1] *Kelmanov A., Mikhailova L., and Romanchenko S.* On a Problem of Summing Elements Chosen from a Family of Finite Numerical Sequences // Lecture Notes in Computer Science, Springer, 2018 (in press).

Улучшенная полиномиальная приближённая схема для задачи маршрутизации транспорта с ограничениями на грузоподъёмность и временные промежутки обслуживания

Хачай Михаил Юрьевич^{1,2,3}

mkhachay@imm.uran.ru

Огородников Юрий Юрьевич^{1,2}

yogorodnikov@gmail.com

¹Екатеринбург, Институт математики и механики им.

Н.Н.Красовского Уро РАН

²Екатеринбург, Уральский Федеральный университет

³Омск, Омский государственный технический университет

DOI: 10.30826/IDP201829

Задача маршрутизации транспорта с ограничениями на грузоподъёмность транспортных средств и временные промежутки обслуживания клиентов (англ. CVRPTW) является широко известной NP-трудной задачей комбинаторной оптимизации. В работе [1] описана полиномиальная приближённая схема, являющаяся улучшением классической схемы М.Хаймовича и А.Ринной-Кана, состоящая в декомпозиции множества обслуживаемых клиентов X на три подмножества X_{out} , X_{mid} и X_{in} . Для клиентов $X_{out} \cup X_{mid}$ строится оптимальное решение U такое, что все клиенты X_{out} обязательно посещаются маршрутами из U , а клиентов X_{mid} посещать не обязательно, однако за пропуск каждого из них накладывается штраф, заключающийся в отдельном посещении данного клиента. Для множества X_{in} применяется эвристика итерированного разбиения маршрутов ИТР. Предложенный алгоритм для произвольного $\varepsilon > 0$ за время $\text{TIME}(\text{TSP}, \rho, n) + O(n^2) + O(e^{O(q(\frac{q}{\varepsilon})^3(p\rho)^2 \log(p\rho))})$ находит $(1 + \varepsilon)$ -приближённое решение задачи CVRPTW на евклидовой плоскости, где q — верхняя оценка грузоподъёмности транспортных средств, p — число промежутков обслуживания (временных окон) и $\text{TIME}(\text{TSP}, \rho, n)$ — трудоёмкость поиска ρ -приближённого решения вспомогательной постановки метрической задачи коммивояжера.

Работа поддержана грантом РФФИ № 14-11-00109.

- [1] *Хачай М. Ю., Огородников Ю. Ю.* Полиномиальная приближённая схема для задачи маршрутизации транспорта с ограничениями по грузоподъёмности и временным окнам // Тр. Ин-та математики и механики УрО РАН. 2018. Т. 24, №3. (в печати).

Improved Polynomial Time Approximation Scheme for Capacitated Vehicle Routing Problem with Time Windows

Khachay Michael^{1,2,3}

mkhachay@imm.uran.ru

Ogorodnikov Iurii^{1,2}

yogorodnikov@gmail.com

¹Krasovsky Institute of Mathematics and Mechanics, Ekaterinburg, Russia

²Ural Federal University, Ekaterinburg, Russia

³Omsk State Technical University, Omsk, Russia

Capacitated Vehicle Routing Problem with Time Windows (CVRPTW) is the well-known combinatorial optimization problem that is NP-complete. In paper [1] the improved polynomial approximation scheme based on M.Haimovich and A.Rinnoy-Kan is proposed. The main idea of this scheme is to decompose a client set X onto three subsets X_{out} , X_{mid} and X_{in} . For clients $X_{out} \cup X_{mid}$ there should be find an optimal solution U that visits all clients from X_{out} but not all from X_{mid} . If any client from X_{mid} is not visited, it should be penal by separate visiting. For clients of X_{in} the iteration partition tour heuristic is applied. So, for any fixed $\varepsilon > 0$, the scheme proposed finds a $(1 + \varepsilon)$ -approximate solution of CVRPTW in time

$$TIME(TSP, \rho, n) + O(n^2) + O\left(e^{O\left(q\left(\frac{q}{\varepsilon}\right)^3(p\rho)^2 \log(p\rho)\right)}\right),$$

for capacity q , number of time windows p to service customers, and time $TIME(TSP, \rho, n)$ needed for finding ρ -approximate solution for an auxiliary instance of the metric TSP. It can be noticed, that the proposed scheme is EPTAS for the fixed capacity $q \geq 1$ and the number of time windows $p \geq 1$ that finds $(1 + \varepsilon)$ -approximate solution for the planar Euclidean CVRPTW in time $O(n^3 + \exp((1/\varepsilon)^3))$. The scheme remains PTAS, when $p^3 q^4 = \log n$.

This research was supported by Russian Science Foundation, grant no. 14-11-00109.

- [1] *Khachay M., Ogorodnikov I.* Improved Polynomial Time Approximation Scheme for Capacitated Vehicle Routing Problem with Time Windows // Trudy Instituta Matematiki i Mekhaniki URO RAN, 2018, vol. 24, no. 3. (in press)

Неулучшаемая гарантированная оценка точности для задачи о k медианах на отрезке

Хачай Михаил Юрьевич^{1,2,*}

daniil.khachay@gmail.com

Хачай Даниил Михайлович^{1,2}

mkhachay@imm.uran.ru

¹Екатеринбург, Институт математики и механики УрО РАН

²Екатеринбург, Уральский Федеральный Университет

DOI: 10.30826/IDP201830

Одномерная задача кластеризации k -medians рассматривается в контексте игры двух лиц с нулевой суммой. Множество стратегий первого игрока совпадает с совокупностью выборок фиксированной длины из отрезка $[0, 1]$. Стратегиями второго игрока являются всевозможные разбиения произвольной выборки данной длины на заданное число кластеров. В качестве платежной выступает функция, оценивающая качество кластеризации, значение которой численно совпадает с суммой уклонений элементов выборки от центров ближайших к ним кластеров. Как нетрудно убедиться, за исключением редких случаев данная игра не имеет цены. Для произвольных натуральных n и k строится верхняя оценка $0.5n/(2k - 1)$ нижней цены игры. Обосновывается достижимость найденной оценки при $k > 1$ и достаточно больших $n = n(k)$. Тем самым показывается, что для произвольной выборки длины n может быть построена кластеризация методом k медиан так, что значение платежной функции не превысит найденной оценки, причем данная оценка достижима при произвольном числе кластеров и выборке достаточно большой длины. Полученные результаты нашли применение в комбинаторной оптимизации при обосновании полиномиальной разрешимости подклассов трудно-решаемых экстремальных задач

Работа поддержана грантом РФФИ № 16-07-00266, 16-01-00505, и 17-08-01385.

- [1] Khachay, M., Khachay, D. Attainable accuracy guarantee for the k -medians clustering in $[0, 1]$ // Optim Lett (2018). Springer, 2018. <https://doi.org/10.1007/s11590-018-1305-3>.

Attainable accuracy guarantee for the k -medians clustering in a line segment

Michael Khachay^{1,2*}

daniil.khachay@gmail.com

Daniel Khachay^{1,2}

mkhachay@imm.uran.ru

¹Ekaterinburg, Krasovsky Institute of Mathematics and Mechanics

²Ekaterinburg, Ural Federal University

We consider the famous k -medians clustering problem in the context of a zero-sum two-player game, which is defined as follows. For given integers $n > 1$ and $k > 1$, strategy sets of the first and second players consist of n -samples drawn from the unit segment $[0, 1]$ and partitions of the index set $\{1, \dots, n\}$ into k nonempty subsets (clusters), respectively. As a payoff, we take a loss function of the k -medians clustering evaluated in terms of the sample chosen by the first player and the partition taken by the second one. Actually, the payoff coincides with the sum of distances between points of the sample and the nearest center of a cluster. It is easy to verify that this game has no value.

In this paper, for any $n > 1$ and $k > 1$, we show that $0.5n/(2k-1)$ is an upper bound for the lower value of this game. Furthermore, for any k , we prove attainability of this bound for some $\bar{n} = \bar{n}(k)$ and an arbitrary $n \geq \bar{n}$. As a consequence, we show that any n -sample from $[0, 1]$ can be partitioned into k clusters, such that the value of k -medians clustering criterion does not exceed the bound obtained and this bound is tight for sufficiently large n .

This research is supported by RFBR, grants no. 16-07-00266, 16-01-00505, and 17-08-01385.

- [1] *Khachay, M., Khachay, D.* Attainable accuracy guarantee for the k -medians clustering in $[0, 1]$ // *Optim Lett* (2018). Springer, 2018. <https://doi.org/10.1007/s11590-018-1305-3>.

Разрешимость евклидовой обобщенной задачи коммивояжера на сетке фиксированной высоты

Хачай Михаил Юрьевич^{1,2}

mkhachay@imm.uran.ru

Незнахина Екатерина Дмитриевна^{1,2}

eneznakhina@yandex.ru

¹Екатеринбург, Институт математики и механики УрО РАН

²Екатеринбург, Уральский федеральный университет

DOI: 10.30826/IDP201831

В работе рассмотрена евклидова обобщенная задача коммивояжера на сетке (Euclidean Generalized Traveling Salesman Problem in Grid Clusters, EGTSP-GC), являющаяся геометрическим подклассом классической обобщенной задачи коммивояжера (Generalized TSP, GTSP), введенной Бхаттачария и соавторами. В их статье показано, что EGTSP-GC NP-трудна в сильном смысле в случае, если количество кластеров k является частью входа, и предложен первый полиномиальный алгоритм с гарантированной оценкой точности. Недавно нам удалось построить для EGTSP-GC PTAS при $k = O(\log n)$ и $k = n - O(\log n)$. Для любого фиксированного k EGTSP-GC может быть решена точно за полиномиальное время как частный случай GTSP. Интересно было бы описать наиболее общую постановку задачи, обладающую этим свойством. Недавно, опираясь на теорию обобщенных пирамидальных маршрутов, мы предложили точный алгоритм с кубической трудоемкостью для постановки задачи EGTSP-GC, в которой высота сетки не превышает 2. Этот результат удалось обобщить на случай сетки произвольной фиксированной высоты. В работе было показано, что с помощью метода динамического программирования произвольная постановка EGTSP-GC, заданная на сетке высоты h , может быть решена точно за время $O(2^{l(h)} n^{l(h)+3})$, где $l(h) = 15h^3 + 2h$.

Работа поддержана грантом РФФ № 14-11-00109.

- [1] *Khachay M., Neznakhina K.* Towards Tractability of the Euclidean Generalized Traveling Salesman Problem in Grid Clusters Defined by a Grid of Bounded Height // In: Ereemeev A., Khachay M., Kochetov Y., Pardalos P. (eds) Optimization Problems and Their Applications, OPTA 2018. Springer, 2018. — P. 68–77.

Towards Tractability of the Euclidean Generalized Traveling Salesman Problem in Grid Clusters Defined by a Grid of Bounded Height

Khachay Michael^{1,2}

mkhachay@imm.uran.ru

Neznakhina Katherine^{1,2}

eneznakhina@yandex.ru

¹Ekaterinburg, Krasovsky Institute of Mathematics and Mechanics

²Ekaterinburg, Ural Federal University

We consider the Euclidean Generalized Traveling Salesman Problem in Grid Clusters (EGTSP-GC), a special geometric subclass of the famous Generalized TSP, introduced by Bhattacharya et al. They showed that the problem is strongly NP-hard if the number of clusters k belongs to the instance and proposed the first polynomial time algorithm with a fixed approximation ratio. Recently, we proved that EGTSP-GC belongs to PTAS when $k = O(\log n)$ and $k = n - O(\log n)$. Meanwhile, being the special case of GTSP, for any fixed k , EGTSP-GC can be solved to optimality in polynomial time. Therefore, it seems interesting to describe the most general case of the problem sharing this property. Recently, by virtue of generalized pyramidal routes, we provided an optimal algorithm with $O(n^3)$ time complexity bound for the case of EGTSP-GC, whose grid height does not exceed 2. We extend this result to the case of EGTSP-GC defined by a grid of any fixed height. It was shown in the paper that any instance of the EGTSP-GC on a grid of height h can be solved to optimality in time $O(2^{l(h)}n^{l(h)+3})$, where $l(h) = 15h^3 + 2h$.

This research was supported by RSF grant 14-01-00109.

- [1] *Khachay M., Neznakhina K.* Towards Tractability of the Euclidean Generalized Traveling Salesman Problem in Grid Clusters Defined by a Grid of Bounded Height // In: Ereemeev A., Khachay M., Kochetov Y., Pardalos P. (eds) Optimization Problems and Their Applications, OPTA 2018. Springer, 2018. — P. 68–77.

Поиск глобального минимума функционала на основе технологий обучения с подкреплением

Зароднюк Татьяна Сергеевна

tz@icc.ru

*Горнов Александр Юрьевич**

gornov@icc.ru

Иркутск, Институт динамики систем и теории управления имени

В.М. Матросова СО РАН

DOI: 10.30826/IDP201832

Основные вычислительные затраты при решении задач оптимального управления в большинстве случаев приходятся на интегрирование систем дифференциальных уравнений при различных «пробных» управлениях. При решении прикладных задач с жесткими управляемыми системами соотношение затрат на вычисление функционалов может достигать нескольких порядков. В таком случае можно трактовать задачи оптимального управления как задачи вычислительно трудоемкие («expensive function optimization problems»). Популярным подходом к этому классу экстремальных проблем в последние годы становится методика Model Based Algorithms («алгоритмы, основанные на моделях»).

В докладе рассматривается эвристический алгоритм поиска экстремума в невыпуклой задаче оптимального управления, основанный на идее накопления информации о значениях функционала и прогнозирования значения на «новом» управлении. Конструкция алгоритма позволяет неявно строить на множестве достижимости системы поверхность уровня оптимизируемого терминального функционала с применением алгоритмов генерации квазислучайных управлений кусочно-линейного или сплайнового типа. В качестве модели, позволяющей аккумулировать уже имеющуюся информацию о функционале, используется функция Шепарда, демонстрирующая хорошие вычислительные характеристики на целом ряде разнородных многомерных задач аппроксимации.

Работа поддержана грантом РФФИ № 17-07-00627.

- [1] *Gornov A. Zarodnyuk T. Anikin A.* The computational technique for nonlinear nonconvex optimal control problems based on a modified gully method // Communications in Computer and Information Science, Springer, 2018 (в печати).

A global search of minimum functional based on reinforcement learning

Zarodnyuk Tatiana

zarodnyuk@icc.ru

*Gornov Alexander**

gornov@icc.ru

Irkutsk, Matrosov Institute for System Dynamics and Control Theory of SB RAS

The main computational costs in solving optimal control problems in most cases are to integrate systems of differential equations for various “trial” controls. When solving applied problems with rigid controllable systems, the ratio of the costs of calculating functionals can reach several orders. In this case, one can treat optimal control problems as “expensive function optimization problems”. A popular approach to this class of extreme problems in recent years is the Model Based Algorithms.

The paper deals with a heuristic algorithm for the extremum search in the nonconvex optimal control problem based on the idea of accumulating information about the values of the functional and predicting the value on the “new” control. The algorithm allows implicitly constructing the surface of the level of the optimized terminal functional on the reachable set of the system, using algorithms for generating quasi-random controls of piecewise linear or spline type. We use Shepard’s function for accumulate already available information about the functional.

This research is funded by RFBR, grant 17-07-00627.

- [1] *Gornov A. Zarodnyuk T. Anikin A.* The computational technique for nonlinear nonconvex optimal control problems based on a modified gully method // *Communications in Computer and Information Science*, Springer, 2018 (in press).

Ускоренный метод для децентрализованной распределённой оптимизации на меняющихся со временем сетях

Рогозин Александр Викторович^{1,*}

aleksandr.rogozin@phystech.edu

*César A. Uribe*²

cauribe2@illinois.edu

Гасников Александр Владимирович^{1,5}

gasnikov@yandex.ru

*Мальковский Николай Владимирович*³

malkovskynv@gmail.com

Angelia Nédich^{1,4}

angelia.nedich@asu.edu

¹Московский физико-технический институт

²University of Illinois at Urbana-Champaign

³Санкт-Петербургский государственный университет

⁴Arizona State University

⁵Институт проблем передачи информации

DOI: 10.30826/IDP201833

Рассмотрим задачу машинного обучения с вектором параметров $y \in \mathbb{R}^d$ и функцией потерь $L(\mathbf{A}, y)$. Пусть, выборка \mathbf{A} не может находиться в памяти одного компьютера из-за своего размера, и поэтому разделена на n частей $\{\mathbf{A}_i\}_{i=1}^n$ и размещена на n машинах.

Задача принимает вид $L(\mathbf{A}, y) = \sum_{i=1}^n L(\mathbf{A}_i, y) \rightarrow \min_{y \in \mathbb{R}^d}$. Это при-

водит к задаче вида $\varphi(y) = \sum_{i=1}^n \varphi_i(y) \rightarrow \min_{y \in \mathbb{R}^d}$, где $\varphi_i : \mathbb{R}^d \rightarrow \mathbb{R}$

является выпуклой для $i = 1, \dots, n$. В работе рассматривается распределённое решение задачи, при котором функции φ_i содержатся локально на отдельных узлах вычислительной сети, а сама сеть меняется во времени из-за технических неисправностей или иных причин. Основным результатом работы является теоретическое обоснование и экспериментальная проверка линейной сходимости ускоренного метода Нестерова для распределённого решения при меняющейся со временем сети. Этот результат получен при определённых предположениях, таких как ограничения на частоту изменения сети вычислителей, а также сильная выпуклость и липшицевость градиента для каждого φ_i .

- [1] *A. Rogozin, C. A. Uribe, A. Gasnikov, N. Malkovsky, A. Nédich.* Optimal distributed convex optimization on slowly time-varying graphs // arXiv, <https://arxiv.org/pdf/1805.06045.pdf>.

An accelerated method for decentralized distributed optimization on time-varying graphs

*Alexander Rogozin*¹★ aleksandr.rogozin@phystech.edu
*César A. Uribe*² cauribe2@illinois.edu
Alexander Gasnikov^{1,5} gasnikov@yandex.ru
*Nikolay Malkovsky*³ malkovskynv@gmail.com
Angelia Nédich^{1,4} angelia.nedich@asu.edu

¹Moscow Institute of Physics and Technology

²University of Illinois at Urbana-Champaign

³Academic University Saint-Petersburg

⁴Arizona State University

⁵Institute for Information Transmission Problems RAS

Suppose one is given a machine learning problem with a vector of parameters $y \in \mathbb{R}^d$ and a loss function $L(\mathbf{A}, y)$. Moreover, assume that the dataset \mathbf{A} is divided into n parts $\{\mathbf{A}_i\}_{i=1}^n$ and placed on n different machines. The corresponding empirical loss minimization problem can be written as

$$L(\mathbf{A}, y) = \sum_{i=1}^n L(\mathbf{A}_i, y) \longrightarrow \min_{y \in \mathbb{R}^d}.$$

This leads to optimization problem of the form

$$\varphi(y) = \sum_{i=1}^n \varphi_i(y) \longrightarrow \min_{y \in \mathbb{R}^d},$$

where $\varphi_i : \mathbb{R}^d \rightarrow \mathbb{R}$ is convex for each $i = 1, \dots, n$. We focus on the distributed problem where each of the functions φ_i is privately held by separate computational units over a network, and the network changes in time due to technical malfunctions or other reasons.

As a result, we theoretically show and empirically verify linear convergence of accelerated Nesterov method. This result is obtained under certain assumptions, which include a constraint on the frequency of the network graph changes and also strong convexity and smoothness of each φ_i .

- [1] *A. Rogozin, C. A. Uribe, A. Gasnikov, N. Malkovsky, A. Nédich*. Optimal distributed convex optimization on slowly time-varying graphs // arXiv, <https://arxiv.org/pdf/1805.06045.pdf>.

Задача распределения ресурсов с точки зрения композитной оптимизации

Иванова Анастасия Сергеевна^{1*}

anastasiya.s.ivanova@phystech.edu

Двуреченский Павел Евгеньевич^{2,3}

pavel.dvurechensky@wias-berlin.de

Гасников Александр Владимирович^{1,3} gasnikov@yandex.ru

¹Москва, Московский физико-технический институт

²Берлин, Институт прикладного анализа и стохастики Вейерштрасса

³Москва, Институт проблем передачи информации РАН

DOI: 10.30826/IDP201834

В данной работе рассматривается задача распределения ресурсов как задача выпуклой минимизации с лиценными ограничениями. Для решения данной задачи были использованы метод проекции субградиента и градиентный спуск, применяемые к двойственной задаче. Была доказана сходимости и получены оценки скорости сходимости как для прямых итераций, так и для двойственных. Так же данные метода были экономически проинтерпретированы. Это означает, что итерации алгоритмов можно соотносить с процессу корректировки цен и производства, чтобы получить желаемый объем производства в экономике. В целом, мы показываем, как эти действия экономических агентов приводят всю систему к равновесию.

Работа А. В. Гасникова и П. Е. Двуреченского поддержана грантом РФФИ № 8-29-03071. Работа А. С. Ивановой поддержана грантом президента Российской Федерации МД-1320.2018.1.

- [1] *Иванова А. С., Двуреченский П. Е., Гасников А. В.* Задача распределения ресурсов с точки зрения композитной оптимизации // *Машинное обучение и анализ данных*, Москва, 2018.

Composite optimization for the resource allocation problem

Ivanova Anastasiya^{1*} anastasiya.s.ivanova@phystech.edu
Dvurechensky Pavel^{2,3} pavel.dvurechensky@wias-berlin.de
Gasnikov Alexander^{1,3} gasnikov@yandex.ru

¹Moscow, Moscow Institute of Physics and Technology

²Berlin, Weierstrass Institute for Applied Analysis and Stochastics

³Moscow, Institute for Information Transmission Problems RAS

In this paper we consider resource allocation problem stated as a convex minimization problem with linear constraints. To solve this problem, we use subgradient method and gradient descent applied to the dual problem and prove the convergence rate both for the primal iterates and the dual iterates. We also provide economic interpretation for these two methods. This means that iterations of the algorithms naturally correspond to the process of price and production adjustment in order to obtain the desired production volume in the economy. Overall, we show how these actions of the economic agents lead the whole system to the equilibrium.

The work of A.V. Gasnikov and P.E. Dvurechensky was supported by RFBR grant 18-29-03071 mk. The work of A.S. Ivanova was supported by Russian president grant MD-1320.2018.1.

[1] *Ivanova A., Dvurechensky P., Gasnikov A.* Composite optimization for the resource allocation problem // *Journal of Machine Learning and Data Analysis*, 2018.

К вопросу существования солитонных решений для систем с полиномиальным потенциалом и их численная реализация

*Бекларян Левон Андреевич*¹*

beklar@cemi.rssi.ru

Бекларян Армен Левонович^{2,1}

abeklaryan@hse.ru

¹Москва, Центральный Экономико-Математический Институт РАН

²Москва, Национальный исследовательский университет «Высшая школа экономики»

DOI: 10.30826/IDP201835

Для уравнений математической физики, являющихся уравнением Эйлера-Лагранжа соответствующих вариационных задач, важный класс решений – это солитонные решения. Изучение солитонных решений основано на существовании взаимно однозначного соответствия между солитонными решениями для исходных систем и решениями индуцированных функционально-дифференциальных уравнений точечного типа (ФДУТТ). Теорема существования и единственности решения для индуцированного ФДУТТ гарантирует существование и единственность солитонного решения с заданными начальными значениями для систем с квазилинейным потенциалом. Для систем с квазилинейным потенциалом также удается сформулировать условия существования периодического решения. Систему с полиномиальным потенциалом можно переопределить так, чтобы возникающий потенциал оказался квазилинейным. Если гарантированное периодическое солитонное решение для такой переопределенной системы будет лежать в шаре, вне которого переопределялся потенциал, то мы получим условия существования периодического солитонного решения для исходной системы с полиномиальным потенциалом. Важной задачей является численная реализация периодических солитонных решений для систем с полиномиальным потенциалом, которая успешно решена.

Исследование частично выполнено за счет гранта Российского научного фонда (проект 17-71-10116), а также работа выполнена при частичной поддержке РФФИ (грант 16-01-00110 А).

- [1] *Beklaryan L. A., Beklaryan A. L.* On the existence of soliton solutions for systems with a polynomial potential and their numerical realization // Машинное обучение и анализ данных, 2018.

On the existence of soliton solutions for systems with a polynomial potential and their numerical realization

*Beklaryan Levon*¹★ beklar@cemi.rssi.ru, lbeklaryan@outlook.com

Beklaryan Armen^{2,1}★ abeklaryan@hse.ru

¹Moscow, Central Economics and Mathematics Institute RAS

²Moscow, National Research University Higher School of Economics

For equations of mathematical physics, which are the Euler-Lagrange equation of the corresponding variational problems, an important class of solutions are soliton solutions. The study of soliton solutions is based on the existence of a one-to-one correspondence between soliton solutions for initial systems and solutions of induced functional-differential equations of pointwise type (FDEPT). The existence and uniqueness theorem for an induced FDEPT guarantees the existence and uniqueness of a soliton solution with given initial values for systems with a quasilinear potential. For systems with a quasilinear potential, one can also formulate the conditions for the existence of a periodic solution. A system with a polynomial potential can be redefined so that the resulting potential turns out to be quasilinear. If a guaranteed periodic soliton solution for such an overdetermined system lies in a sphere, outside which the potential is redefined, then we obtain the conditions for the existence of a periodic soliton solution for the initial system with a polynomial potential. An important task is the numerical realization of periodic soliton solutions for systems with a polynomial potential, which has been successfully solved.

The reported study was partially supported by Russian Science Foundation, Project 17-71-10116. Also this work was partially funded by RFBR according to the research project 16-01-00110 A.

- [1] *Beklaryan L. A., Beklaryan A. L.* On the existence of soliton solutions for systems with a polynomial potential and their numerical realization // Machine Learning and Data Analysis, 2018.

Идентификация сигналов на основе максимально правдоподобного соответствия их случайной выборки стохастическим описаниям прецедентов

Анциперов Вячеслав Евгеньевич^{1,2} antciperov@cplire.ru

¹Москва, ИРЭ им. В.А.Котельникова РАН

²Москва, Московский физико-технический институт

DOI: 10.30826/IDP201836

В докладе обсуждаются результаты синтеза и анализ качества работы нового метода идентификации сигналов по форме их огибающих. Под идентификацией понимается соотнесение формы огибающей тестируемого сигнала с формой огибающей одного из ранее наблюдаемых прецедентов. При этом подразумевается, что исходными данными регистрации как тестируемого сигнала, так и имеющихся прецедентов являются выборки их дискретных отсчетов. Предлагаемая постановка задачи мотивирована особенностями восприятия внешних сигналов-стимулов живыми системами, в частности, человеком. Синтез предлагаемого метода удалось осуществить на основе стат. методов теории точечных процессов, формализовав рассматриваемую проблему в задачу о идентификации оцениваемой по реализации случайных точек формы интенсивности пуассоновских точечных процессов [1]. Ориентируясь с самого начала на методы, применяемые в сфере машинного обучения, в контексте выбранного формализма поставлены и решены проблемы выбора дескрипторов, составляющих описание прецедентов, и задания меры соответствия, подобия в терминах выбранных описаний интенсивностей идентифицируемых процессов и прецедентов. Для выбранной меры подобия разработана эффективная процедура вычисления, близкая по структуре вариационно-байесовским выводам. Процедура доведена до алгоритмической реализации, интерпретации алгоритмов в духе алгоритмов семейства (VB) EM посвящена заключительная часть доклада. Работа поддержана грантом РФФИ № 18-07-01295.

- [1] *Анциперов В.* Определение формы интенсивности точечного процесса с помощью максимально правдоподобных распределений отсчетов прецедентов //Журнал радиоэлектроники, М.: ИРЭ им. В.А.Котельникова, N 12, 2017. <http://jre.cplire.ru/jre/dec17/>.

Signals identification based on the maximum–likelihood matching of their random samples to available precedents stochastic descriptions

Antsiperov Viacheslav^{1,2}

antsiperov@cplire.ru

¹Moscow, Kotelnikov Institute of Radioengineering of RAS

²Moscow, Moscow Institute of Physics and Technology

The report is devoted to the synthesis and quality analysis of a new method for identifying signals by the shape of their envelopes. By identification is meant the comparison of the shape of the tested signal envelope with the shape of one of the previously observed precedents. It is implied that the registered data of both the tested signal and the available precedents are initially represented as their discrete samples. The statement of the problem proposed is motivated by the mechanisms of signals – stimuli perception by living systems, in particular, by man.

The synthesis of the method proposed was realized on the basis of the statistical methods of the theory of point processes, after formalizing the problem in question in the problem of identifying the form of intensity of Poisson point processes by the realization of its random points [1]. Focusing from the very beginning on the methods used in the field of machine learning, two main problems were formulated and solved: the problem of selecting descriptors, which are the basis for precedent description, and the problem of choosing a measure of conformity, similarity of identifiable processes and precedents in terms of selected descriptions.

For a selected similarity measure, an effective calculation procedure is developed that is close in structure to variational Bayesian inference. The procedure is brought to the algorithmic realization. The final part of the report is devoted to the interpretation of developed algorithms in the spirit of the (VB) EM approach. This research is funded by RFBR, grant 18-07-01295.

- [1] *Antsiperov V.* Identification of the point process intensity shape with the precedents maximum likelihood distributions // Zhurnal Radioelektroniki – Journal of Radio Electronics, M.: Kotelnikov Institute of Radioengineering of RAS, N 12, 2017. <http://jre.cplire.ru/jre/dec17/>.

Поиск плавно меняющихся пространственных закономерностей

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

¹Москва, НИУ ВШЭ

²Москва, НИЯУ МИФИ

DOI: 10.30826/IDP201837

В настоящее время бурное развитие геоинформатики и большие объёмы данных, имеющих привязку к местности, позволяют применять методы интеллектуального анализа данных к поиску закономерностей в пространственных данных. В настоящей работе предпринята попытка адаптировать алгоритмы, разрабатываемые авторами для временных рядов, к поиску закономерностей в пространственных данных. Таким образом авторы исследуют закономерности, плавно меняющиеся в пространстве.

Для иллюстрации подхода в работе использованы данные Сбербанка России о стоимости квартир в Москве, рассмотренные авторами в докладе на конференции ММРО-18. Однако в отличие от предыдущей работы, акцент делается не на временной, а пространственной составляющей данных. Анализируется зависимость целевой переменной (стоимости квадратного метра жилья) от района, а также близости к центру города, станциям общественного транспорта, крупным магистралям, магазинам, спортивным и развлекательным центрам, институтам здравоохранения, образовательным учреждениям, офисам, паркам и т.д. Отметим, что предложенный в работе подход к анализу пространственных данных, широко применим и в других областях, например, анализу различных данных со спутников, геотаргетированному маркетингу и т.п.

Работа поддержана грантом РФФИ № 16-07-01156.

- [1] *Филипенков Н. В., Петрова М. А.* Поиск плавно меняющихся пространственных закономерностей // Машинное обучение и анализ данных, Москва: Издательство, 2018.

Mining the Slightly Changing Patterns in Geospatial Data

Nikolay Filipenkov^{1*}

n.filipenkov@mail.ru

*Marina Petrova*²

marina_petrova@mail.ru

¹Moscow, HSE

²Moscow, MEPhI

The rapid growth of geospatial data in the world enables the implementation of data mining techniques to mine the patterns in geospatial data. In this paper the authors have modified the algorithms for mining slightly changing patterns in time series and adapted them to work with geospatial data. So the paper discusses mining the patterns that slightly change in space (instead of time).

The paper uses the same data on the property prices in Moscow provided by Sberbank of Russia that has been discussed in MMPR-18 conference paper by the authors. However in this paper the major focus is made on the space dimension of the data compared to the focus on time in the previous paper. The predicted variable (square meter price) is analyzed respective to the district, distance to the city center, stations of public transport, highways, shops, sports, entertainment, healthcare, education centers, offices, parks etc. The proposed approach for mining slightly changing patterns in geospatial data is highly applicable to any data with geo-tag, e.g. space image recognition, geo-targeted marketing etc.

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

- [1] *Filipenkov N. V., Petrova M. A.* Mining the Slightly Changing Patterns in Geospatial Data // Machine Learning and Data Analysis, Moscow: Publisher, 2018.

Снижение размерности с помощью проекции на скрытое пространство в задаче декодирования сигналов

*Исаченко Роман Владимирович*¹
Стрижов Вадим Викторович^{1,2*}

isa-ro@yandex.ru

strijov@ccas.com

¹Москва, Московский физико-технический институт

²Москва, Вычислительного центра им. А.А. Дородницына

ФИЦ ИУ РАН

DOI: 10.30826/IDP201838

Работа посвящена снижению размерности в задаче декодирования сигнала. Исходное пространство обладает высокой размерностью и является избыточным. В работе рассматривается случай векторной целевой переменной. Корреляции наблюдаются как во входном пространстве объектов так и в пространстве целевой переменной. Для построения простой и устойчивой модели используются методы снижения размерности и выбора признаков. Модель PLS используется в качестве базовой модели. Модель проецирует входные и выходные объекты в скрытое пространство и максимизирует ковариацию между проекциями. Для получения разреженной модели применяются методы выбора признаков. Большинство методов выбора признаков не учитывают зависимости в целевом пространстве. В работе предлагается новый подход к выбору признаков в задаче многомерной регрессии. Предложенный подход обобщает метод выбора признаков с помощью квадратичного программирования (QPFS). Алгоритм QPFS выбирает некоррелированные признаки, релевантные целевой переменной. Предлагаемые методы учитывают зависимости в целевом пространстве и отбирают информативные признаки. Вычислительный эксперимент проводился на наборе данных электрокортикограмм. Предложенные алгоритмы превзошли результаты базовых моделей. Лучший результат получен комбинацией алгоритмов QPFS и PLS. Работа поддержана грантом РФФИ № 16-07-01155 и Правительства РФ, соглашение 05.Y09.21.0018.

- [1] *Исаченко Р. В., Стрижов В. В.* Снижение размерности с помощью проекции на скрытое пространство в задаче декодирования сигналов // <http://www.mathnet.ru/links/bfec367d16826750ba3d8bfadd51c3d8/ia415.pdf>.

Dimensionality Reduction for Multicorrelated Signal Decoding with Projections to Latent Space

*Isachenko Roman*¹

roman.isachenko@phystech.edu

Strijov Vadim^{1,2*}

strijov@ccas.com

¹Moscow, Moscow Institute of Physics and Technology

²Moscow, FRC CSC RAS

The study is devoted to the problem of dimensionality reduction for signal decoding. The challenge of the investigation is redundancy in data description. High correlation in measurements leads to correlation in input space. The study considers multivariate problem, the target variable is a vector. In this case the correlations occur in both input and target spaces. Dimensionality reduction and feature selection are used to build simple and stable model.

Projection to latent space (PLS) regression is used as a base model for the dimensionality reduction. The model projects input and target data into the joint latent space and maximizes the covariances between the projections. To obtain the sparse model the feature selection is applied. The majority of feature selection methods ignore the dependencies in the target space. The study suggests a novel approach to feature selection in multivariate regression. The proposed approach extends the ideas of the quadratic programming feature selection (QPFS) algorithm. The QPFS algorithm selects non-correlated features, which are relevant to the targets. The proposed methods take into account the dependencies in the target space and select features which are informative to all targets jointly.

The computational experiment was carried out on the electrocorticogram dataset. The proposed algorithms were compared by different criteria such as stability and their prediction performance. The algorithms give significantly better results compared to the baseline strategy. The best result is obtained by combination of QPFS and PLS algorithms.

The research was supported by the RFBR, project 16-07-01155 and by the Government of the Russian Federation, agreement 05.Y09.21.0018.

- [1] *Isachenko R., Strijov V.* Dimensionality Reduction for Multicorrelated Signal Decoding with Projections to Latent Space // <http://www.mathnet.ru/links/bfec367d16826750ba3d8bfadd51c3d8/ia415.pdf>.

Моделирование и анализ природных временных рядов на основе обобщенной многокомпонентной модели

Мандрикова Оксана Викторовна

oksanam1@mail.ru

Фетисова Надежда Владимировна

nv.glushkova@ya.ru

Полозов Юрий Александрович

up_agent@mail.ru

Петропавловск-Камчатский, Институт космофизических

исследований и распространения радиоволн ДВО РАН

DOI: 10.30826/IDP201839

Работа направлена на построение методов моделирования и анализа природных временных рядов и создание автоматизированных систем на их основе. Предложена обобщенная многокомпонентная модель природного временного ряда сложной структуры (ОМКМ), позволяющая описать иррегулярные вариации данных. Рекуррентная составляющая модели имеет параметрический вид и описывает регулярный временной ход данных. Аномальные компоненты модели имеют вид нелинейных аппроксимирующих схем и описывают иррегулярные изменения. На примере временных рядов критической частоты F2-слоя ионосферы (использовались данные мировой сети ионосферных станций) описана реализация модели и показаны результаты ее применения. Приведено сравнение с международной эмпирической моделью ИРИ и медианным методом, подтвердившее эффективность ОМКМ. Предлагаемая ОМКМ в отличие от аналогов позволяет в автоматическом режиме выделить аномальные изменения данных и оценить их характеристики. Модель реализована численно и доступна через сеть Интернет (<http://aurorsa.ikir.ru:8580>). Результаты исследования важны в задачах геофизического мониторинга и оперативного прогноза космической погоды.

Исследования выполнены при поддержке Гранта РНФ №14-11-00194.

- [1] *Мандрикова О. В., Фетисова Н. В., Полозов Ю. А.* Моделирование и анализ природных временных рядов на основе обобщенной многокомпонентной модели // Машинное обучение и анализ данных, Москва: Вычислительный центр им. А.А. Дородницына РАН, 2018. — С. 74–88. <http://jmla.org/papers/doc/2018/no2/Mandrikova2018Mcm.pdf>.

Modeling and analysis of natural time series on the basis of general multicomponent model

Mandrikova Oksana

oksanam1@mail.ru

Fetisova Nadezhda

nv.glushkova@ya.ru

Polozov Yuri

up_agent@mail.ru

Petropavlovsk-Kamchatskiy, Institute of Cosmophysical Research and Radio Wave Propagation FEB RAS

The work is focused on the development of methods for modeling and analysis of natural time series and the construction of automated systems on their basis. The present paper proposes a general multicomponent model (GMCM) of complex time series that allows describing irregular variations in the data. The GMCM recurrent component is represented by a parametric form and describes the regular time course of the data. The GMCM anomalous components are represented by nonlinear approximating schemes and describe irregular variations. On the example of the ionospheric critical frequency time series of F2-layer (data from the world network of ionospheric stations were used), the implementation of the model is described, and the results of its application are presented. A comparison with the IRI international empirical model and the median method confirmed the efficiency of the GMCM. The proposed GMCM, in contrast to analogs, allows us to detect anomalous changes in the data and to estimate their characteristics in automatic mode. The model is implemented numerically and is available from the Internet (<http://aurorasa.ikir.ru:8580>). The results of the research are important in the tasks of geophysical monitoring and operational forecast of space weather.

The investigation is supported by RSF Grant No14-11-00194.

- [1] *Mandrikova O, Fetisova N., Polozov Yu.* Modeling and analysis of natural time series on the basis of general multicomponent model // Machine Learning and Data Analysis, Moscow: Computing Centre of A.A. Dorodnicyn RAS, 2018. — p. 74–88. <http://jmla.org/papers/doc/2018/no1/Author2018Title.pdf>.

Анализ данных нейтронных мониторов и выделение спорадических эффектов на основе нейронных сетей и вейвлет-преобразования

*Мандрикова Оксана Викторовна*¹ oksanam1@mail.ru

*Залев Тимур Ленарович*¹ tim.aka.geralt@mail.ru

*Геппенер Владимир Владимирович*² geppener@mail.ru

*Мандрикова Богдана Сергеевна*¹ 555bs5@mail.ru

¹Петропавловск-Камчатский, ИКИР ДВО РАН

²Санкт-Петербург, СПбГЭТУ «ЛЭТИ»

DOI: 10.30826/IDP201840

В работе предложен метод и вычислительные алгоритмы анализа данных космических лучей и выделения спорадических эффектов. Метод основан на применении нейронных сетей и вейвлет-преобразования. Используются нейронные сети векторного квантования и многослойный персептрон. На основе метода по данным мировой сети нейтронных мониторов изучена динамика вариаций космических лучей и определены признаки возникновения спорадических эффектов. Показана эффективность применения нейронных сетей векторного квантования для задачи классификации данных нейтронных мониторов. Приведен способ аппроксимации временного хода космических лучей на основе многослойного персептрона. Описан вычислительный алгоритм детального анализа данных нейтронных мониторов и выделения разномасштабных спорадических эффектов, основанный на непрерывном вейвлет-преобразовании (алгоритм реализован численно (<http://aurorasa.ikir.ru:8580>)). Предложенный метод позволяет в автоматическом режиме выполнять классификацию данных нейтронных мониторов и выделять спорадические эффекты различной продолжительности и интенсивности. Результаты исследования представляют интерес в задачах прогноза комической погоды и предсказания магнитных бурь. Исследование выполнено при финансовой поддержке гранта РФФИ No 14-11-00194-П.

- [1] *Мандрикова О. В., Залев Т. Л., Геппенер В. В., Мандрикова Б. С.* Анализ данных нейтронных мониторов и выделение спорадических эффектов на основе нейронных сетей и вейвлет-преобразования // Машинное обучение и анализ данных, Москва: Вычислительный центр им. А.А. Дородницына РАН, 2018.

Modeling and analysis of natural time series on the basis of general multicomponent model

*Mandrikova Oksana*¹

oksanam1@mail.ru

*Zalyaev Timur*¹

tim.aka.geralt@mail.ru

*Geppener Vladimir*²

geppener@mail.ru

*Mandrikova Bogdana*¹

555bs5@mail.ru

¹Petropavlovsk-Kamchatskiy, IKIR FEB RAS

²St. Petersburg, ETU «LETI»

The work is aimed at developing of automated methods for analysis of cosmic ray variations. The present paper proposes a method and computational algorithms for analysis of cosmic ray data and detection of sporadic effects. The method is based on the use of the neural networks and the wavelet transform. The neural networks of vector quantization and multilayer perceptron are used. Using the proposed method, we studied the dynamics of cosmic ray variations and identified the signs of the occurrence of sporadic effects (the data from the world network of neutron monitors were used). The efficiency of the application of the neural networks of vector quantization for the problem of classification of neutron monitor data in automatic mode is shown. A method for approximating of the cosmic ray time course is presented. The method is based on the neural network of a multilayer perceptron and the fast wavelet transform. A computational algorithm for the detailed analysis of neutron monitor data and detection of multiscale sporadic effects is described. The algorithm is based on the continuous wavelet transform (the algorithm is implemented numerically). The proposed method, in contrast to analogs, allows us to classify neutron monitor data and to detection sporadic effects of various duration and intensity in automatic mode. The results of the research are of interest in the tasks of the forecasting of space weather and predicting of magnetic storms. The investigation is supported by RSF Grant No14-11-00194.

- [1] *Mandrikova O, Zalyaev T., Geppener V., Mandrikova B.* Analysis of neutron monitor data and detection of sporadic effects on the basis of the neural networks and the wavelet-transform // Machine Learning and Data Analysis, Moscow: Computing Centre of. A.A. Dorodnicyn RAS, 2018.

Метод анализа псевдофазового портрета при распознавании биосигналов с нелинейными свойствами

Манило Людмила Алексеевна

lmanilo@yandex.ru

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

DOI: 10.30826/IDP201841

Рассмотрен метод распознавания биомедицинских сигналов с нелинейными свойствами, основанный на реконструкции фазового портрета анализируемой временной последовательности. Исследованы возможности оценки корреляционной размерности временного ряда и размерности пространства вложения с использованием алгоритма Грассбергера-Прокаччия (P. Grassberger, I. Procaccia, 1983). Показано, что классификацию сигналов можно реализовать на основе более простых описаний, например, представлении псевдофазового портрета в пространстве меньшей размерности. Основное внимание уделено морфологическому анализу строящейся фазовой траектории и выбору признаков для распознавания биосигналов с хаотическими свойствами. Приведены результаты исследований, направленных на распознавание по ЭКГ аритмий, отличающихся значительным разбросом и нерегулярным характером изменения длительностей кардиоциклов. В ходе экспериментов установлено, что задачу автоматического обнаружения фрагментов фибрилляции предсердий можно эффективно решить, используя линейный дискриминантный анализ морфологического описания псевдофазового портрета ритмограммы на плоскости смежных кардиоциклов. Обоснован выбор наиболее информативных признаков, предложен алгоритм классификации аритмий по ритмограмме. Результаты исследований могут быть использованы при решении задач кардиологического наблюдения.

Работа поддержана грантами РФФИ № 16-01-00159, 18-29-02036.

- [1] *Хачатрян К. С., Манило Л. А.* Метод анализа псевдофазового портрета в задаче распознавания биомедицинских сигналов // *Биотехносфера*, 2016, 5 (47). — С. 14–18.

The method of analysis of pseudo-phase portrait in recognition of biosignals with nonlinear properties

Manilo Ludmila

lmanilo@yandex.ru

Saint Petersburg, St. Petersburg Electrotechnical University "LETI"

The method of recognition of biomedical signals with nonlinear properties based on the reconstruction of the phase portrait of the analyzed time sequence is considered. The possibilities of estimating the correlation dimension of the time series and the dimension of the embedding space are investigated using the Grassberger-Procaccia algorithm (P. Grassberger, I. Procaccia, 1983). It is shown that the classification of signals can be realized on the basis of simpler descriptions, for example, on the representation of a pseudo-phase portrait in a space of smaller dimension. The main attention is paid to the morphological analysis of the phase trajectory under construction and the choice of features for recognition of biosignals with chaotic properties. The results of studies aimed at the arrhythmias recognition by ECG, characterized by a significant spread and irregular changes in the duration of cardiocycles are presented. During the experiments it was established that the problem of automatic detection of atrial fibrillation fragments can be effectively solved using a linear discriminant analysis of the morphological description of the pseudophase portrait of the rhythmogram on the plane of adjacent cardiac cycles. The choice of the most informative features is justified, the algorithm of arrhythmia classification according to the rhythmogram is proposed. The results of the research can be used in solving the problems of cardiac monitoring.

This research is funded by RFBR, grants 16-01-00159, 18-29-02036.

- [1] *Khachatryan K. S., Manilo L. A.* The method of analysis pseudophase portrait in the problem of recognition of biomedical signals // *Biotechnology*. - 2016, No 5 (47). — p. 14–18.

Оптимизация параллельного алгоритма глобального выравнивания протяженных последовательностей

Панкратов Антон Николаевич^{1,2,*} pan@impb.ru

*Тетюев Руслан Курманбиевич*¹ ruslan.tetuev@gmail.com

*Пятков Максим Иванович*¹ mpyatkov@gmail.com

¹Пушино, Институт математических проблем биологии РАН – филиал Института прикладной математики им.М.В.Келдыша РАН

²Пушино, Пушинский государственный естественно-научный институт

DOI: 10.30826/IDP201842

Цель этой работы заключается в сборе лучших практик и построении наиболее общей модели парного глобального выравнивания с возможностью выравнивания как можно более длинных последовательностей. В результате универсальный алгоритм глобального выравнивания построен на основе параллельной реализации FastLSA алгоритма Needleman-Wunsch'a с произвольной матрицей подстановок и алгоритмом Gotoh'a для аффинной системы штрафов за пробелы.

При построении алгоритма используется стратегия отказа от хранения промежуточных данных в пользу повторяющихся вычислений с открывающейся возможностью распараллеливания вычислений и минимизации использования памяти. Найдены аналитически оптимальные значения параметров распараллеливания метода. Алгоритм реализован на языке программирования Javascript как веб-приложение. Распределенные вычисления реализованы на языке программирования PHP и производятся на клиентах, которые используют данный веб-сервис.

Алгоритм может использоваться при выравнивании текстовых последовательностей в произвольном алфавите.

Работа поддержана грантом РФФИ № 16-01-00692.

- [1] *Тетюев Р. К., Пятков М. И., Панкратов А. Н.* Параллельный алгоритм глобального выравнивания протяжённых аминокислотных и нуклеотидных последовательностей // *Мат. биол. и биоинф.*, 2017.12(1). — С. 137–150. http://www.matbio.org/2017/Tetuev_12_137.pdf.

Optimization of parallel algorithm for global alignment of long sequences

Pankratov Anton^{1,2*}

pan@impb.ru

*Tetuev Ruslan*¹

ruslan.tetuev@gmail.com

*Pyatkov Maxim*¹

mpyatkov@gmail.com

¹Pushchino, Institute of Mathematical Problems of Biology RAS – the branch of Keldysh Institute of Applied Mathematics RAS

²Pushchino, Pushchino State Natural Science Institute

The goal of this work is to collect the best practices and to build the most general model of a pairwise global alignment with the possibility of aligning as long sequences as possible. As a result, the versatile algorithm for global alignment is constructed on the basis of parallel implementation FastLSA of the Needleman-Wunsch algorithm with an arbitrary matrix of substitutions and Gotoh algorithm for the affine system of penalties for gaps.

When constructing an algorithm, the strategy is used to avoid storing intermediate data in favor of repetitive calculations with the opening possibility of parallelizing computations and minimizing the use of memory. The optimal values of parameters of parallelization of the method are found analytically. The algorithm is implemented in the Javascript programming language as web application. Distributed computing is implemented in the PHP programming language and is produced on clients that use this web service.

The algorithm can be used when aligning text sequences in an arbitrary alphabet.

This research is funded by RFBR, grant 16-01-00692.

- [1] *Tetuev R., Pyatkov M., Pankratov A.* Parallel algorithm for global alignment of long aminoacid and nucleotide sequences // *Math. Biol. Bioinf.*, 2017.12(1). — P.137–150. http://www.matbio.org/2017/Tetuev_12_137.pdf.

Спектральные и пространственные особенности энцефалограмм при психических расстройствах

Панкратова Наталья Михайловна^{1*} natpan1974@mail.ru

*Рыкунов Станислав Дмитриевич*¹ stanislavrykunov@gmail.com

*Бойко Анна Ивановна*¹ a.boyko@list.ru

*Молчанова Дина Альбертовна*¹ dm@pochtamt.ru

*Устинин Михаил Николаевич*¹ u_m_n@mail.ru

¹Пуццино, ИМПБ РАН - филиал ИПМ им.М.В.Келдыша РАН

DOI: 10.30826/IDP201843

Рассматривается вопрос изменения энцефалограмм при психических расстройствах. Представленный обзор литературы показывает, что электрическая активность головного мозга в этих случаях имеет как спектральные особенности, так и особенности локализации источников патологической активности в разных частотных диапазонах. Вопрос о пространственном расположении источников патологической активности является ключевым при изучении работы мозга и решается с помощью различных методов локализации. Предложен метод для точного количественного анализа активности по данным энцефалографии. Результаты локализации отображаются на магнитно-резонансной томограмме субъекта, на основе чего строятся гипотезы о нейрофизиологическом механизме изучаемой патологии. Метод, предложенный в данной работе, опирается на преобразование Фурье многоканальных данных энцефалографии и локализацию отдельных спектральных компонент. Это позволяет детально изучать те или иные частотные признаки патологической активности мозга и отвечать на вопросы об их связи с анатомией мозга.

Работа выполнена при поддержке Программы Президиума РАН №27 «Фундаментальные проблемы решения сложных практических задач с помощью суперкомпьютеров» и грантов РФФИ 17-29-02178, 18-29-02087, 16-07-00937, 16-07-01000, 17-07-00677, 17-07-00686.

- [1] *Панкратова Н. М., Рыкунов С. Д., Бойко А. И., Молчанова Д. А., Устинин М. Н.* Локализация спектральных особенностей энцефалограмм при психических расстройствах // Математическая биология и биоинформатика, 2018. Т. 13. № 2. С. 322–336.

doi:10.17537/2018.13.322.

Spectral and spatial features of encephalograms in mental disorders

Pankratova Natalia^{1*}

natpan1974@mail.ru

*Rykunov Stanislav*¹

stanislavrykunov@gmail.com

*Boyko Anna*¹

a.boyko@list.ru

*Molchanova Dina*¹

dm@pochtamt.ru

*Ustinin Mikhail*¹

coauthor@site.ru

¹Pushchino, IMPB RAS- Branch of KIAM RAS

The issue of changing encephalograms in mental disorders is considered. The presented review of the literature shows that the electrical activity of the brain in these cases has both spectral features and features of localization of pathological activity sources in different frequency ranges. Spatial position of the sources of pathological activity is a key issue of brain studies and it is solved by various localization methods. A method is proposed for an accurate quantitative analysis of activity from encephalography data. Localization results are displayed on the magnetic resonance tomogram of the subject, on the basis of which hypotheses about the neurophysiological mechanism of the pathology under study are constructed. The method, proposed in this paper, is based on Fourier transform of multichannel encephalography data and on the localization of spectral components. Such approach permits to study in detail some or other frequency features of the brain pathological activity and to reveal their connections with the brain anatomy.

This research is funded by Programs of the Presidium of the Russian Academy of Sciences No.27 «Fundamental problems of solving complex practical problems with the help of supercomputers» and RFBR, grants 17-29-02178, 18-29-02087, 16-07-00937, 16-07-01000, 17-07-00677, 17-07-00686.

- [1] *Pankratova N., Rykunov S., Boyko A., Molchanova D., Ustinin M.* Localization of Encephalogram Spectral Features in Psychic Disorders // *Mathematical Biology and Bioinformatics*, 2018;13(2):322-336. doi: 10.17537/2018.13.322..

Верификация регрессионных моделей во временных рядах

Кириллюк Игорь Леонидович^{1*}

igokir@rambler.ru

*Сенько Олег Валентинович*²

senkoov@mail.ru

¹Москва, Институт экономики РАН

²Москва, Федеральный государственный центр “Информатика и управление” РАН

DOI: 10.30826/IDP201844

Многие прикладные задачи эконометрики связаны с анализом не стационарных временных рядов. Есть различные типы нестационарности. Например, нестационарность может быть связана с реально существующими временными тенденциями (трендами, циклами и т.п.). Также она может быть связана с тем, что изменения во временном ряду имеют характер типа броуновского движения. Практика и эксперименты с сгенерированными данными показывают, что при нестационарности последнего типа возникают эффекты ложной регрессии. Под ложной регрессией понимаются случайно возникшие регрессионные зависимости, которые при использовании стандартных средств верификации формально являются статистически достоверными. Целью проведённых нами исследований на основе временных рядов, включающих значения логарифмов занятости, основных фондов и валового регионального продукта для 79 российских регионов являлось изучение вероятности возникновения формально достоверных ложных регрессий, а также возможности верификации таких регрессий с использованием технологии бутстреп-анализа. Эксперименты проводились путем генерации исходных выборок, очищенных от детерминированных трендов, новых случайных выборок, где приросты значений переменных выбирались из множеств исходных приростов с возвращением. Эксперименты показали высокую вероятность появления регрессии с $R^2 > 0.5$. Вероятность появления таких зависимостей оценивается на уровне менее 0.0001 исходя из гипотезы о независимости наблюдений.

- [1] *Kirilyuk I. L., Kuznetsova A. V., Senko O. V.* Data Mining in Institutional Economics Tasks // EPJ Web Conf., 173 03013, 2018. — p.1–4. https://www.epj-conferences.org/articles/epjconf/pdf/2018/08/epjconf_mmcp2018_03013.pdf.

Verification of regression models in time series

*Kirilyuk Igor*¹

igokir@rambler.ru

*Senko Oleg*²★

senkoov@mail.ru

¹Moscow, Institute of Economics of RAS

²Moscow, FRC "Informatics and Control" of RAS

Many applied problems of econometrics are connected with the analysis of non-stationary time series. There are different types of non-stationarity. For example, nonstationarity can be associated with real time tendencies (trends, cycles, etc.). It can also be related to the fact that changes in the time series have a character like the Brownian motion.

Practice and experiments with artificially generated data show that non-stationarity of latter type may cause so called spurious regression effect. Spurious regressions are randomly arising regression dependences that are formally proved to be statistically valid if standard verification tools are used. Our researches were made on the basis of the time series including values of logarithms of employment, fixed assets and a gross regional product for 79 regions of Russia. The goal of the studies was to evaluate probability of spurious regressions emergence and also assess if bootstrap-analysis technique may be used for their correct verification. Experiments were made by generation of new random samples from the initial samples cleaned from the determined trends. At that increments of variables values were chosen with returns from initial samples. The experiments showed a high probability of regression with $R^2 > 0.5$. The probability of such regression must be less 0.0001 if they are verified with the help of method based on hypothesis about independence of observations.

- [1] *Kirilyuk I., Kuznetsova A., Senko O.* Data Mining in Institutional Economics Tasks // EPJ Web Conf., 173 03013, 2018. — p. 1–4. https://www.epj-conferences.org/articles/epjconf/pdf/2018/08/epjconf_mmcp2018_03013.pdf.

Оценивание состава инвестиционного портфеля в большом множестве биржевых активов

*Моттль Вадим Вячеславович*¹ vmottl@yandex.ru

*Красоткина Ольга Вячеслав.*² okrasotkina@markovprocesses.com

*Морозов Алексей Олегович*³ ao.morozov@phystech.edu

¹Москва, ФИЦ «Информатика и управление» РАН

²Summit, NJ, USA, Markov Processes International

³Москва, Московский физико-технический институт

DOI: 10.30826/IDP201845

Методология восстановления скрытого состава инвестиционного портфеля по наблюдаемому временному ряду его периодических доходностей известна в эконометрике как Returns Based Style Analysis. Ее создатель Уильям Шарп сформулировал задачу анализа данных как задачу регрессионного анализа в предположении, что множество классов ценных бумаг, среди которых ищется распределение капитала, известно и мало по сравнению с числом наблюдений. Дополняя постановку Шарпа мы допускаем, что неизвестная совокупность ценных бумаг, составляющих исследуемый инвестиционный портфель, является лишь малым подмножеством среди обращающихся на рынке биржевых активов, число которых многократно превосходит число наблюдений. Очевидно, что такая регрессионная задача некорректна. Мы используем в качестве регуляризатора естественное предположение, что искомый портфель сформирован с целью сохранения и умножения капитала, что позволяет фундаментально сократить множество рассматриваемых вариантов его распределения между активами. Математическим инструментом реализации этого предположения служит понятие эффективного портфеля, сформулированное Гарри Марковицем, лауреатом той же нобелевской премии. Регрессионный анализ доходностей исследуемого портфеля и биржевых активов сводится к определению одного параметра, называемого Risk Tolerance – склонность инвестора к риску. Работа поддержана грантами РФФИ 17-07-00436, 17-07-00993.

- [1] *O. Krasotkina, M. Markov, V. Mottl, D. Babichev, I. Pugach, A. Morozov.* Constrained Regularized Regression Model Search in Large Sets of Regressors // Lecture Notes in Artificial Intelligence, Vol. 1035, Springer, 2018, pp. 1-15.

Factor Search in Returns Based Analysis of Investment Portfolios

*Vadim Mottl*¹*

vmottl@yandex.ru

*Olga Krasotkina*²

okrasotkina@markovprocesses.com

*Alexey Morozov*³

ao.morozov@phystech.edu

¹Moscow, Computer Science and Control, Russian Academy of Sciences

²Summit, NJ, USA, Markov Processes International

³Moscow Institute of Physics and Technology

The methodology of recovering the hidden sharing of the capital in the portfolio of an investment company from the time series of periodic returns is commonly adopted in econometrics as Returns Based Style Analysis. Its author, Nobel Prize winner William Sharpe, formulated the respective problem as that of regression analysis under the assumption of the known set of securities, among which the capital sharing is sought for, is small in comparison with the number of observed holding periods. As an extension of Sharpe's problem statement, we admit that the unknown set of securities in the investment portfolio under analysis is not known in advance and makes a small subset in the universe of stock market assets, whose number far exceeds the length of the registered returns time series. It is obvious that such a regression problem is ill posed. We issue from the natural assumption that the sought-for portfolio is composed with the purpose to retain and enlarge the capital, which suggestion results in the possibility to fundamentally shrink the variety of capital sharing versions. We harness the notion of an effective portfolio formulated by Harry Markowitz, winner of the same Nobel prize, as the mathematical instrument of such a regularization. The computer assisted analysis of the returns time series on both portfolio under study and stock market assets boils down to finding only one parameter of the regression model, which has the sense of Risk Tolerance – propensity of the investor to risk.

This research is funded by RFBR, grants 17-07-00436 and 17-07-00993.

- [1] *O. Krasotkina, M. Markov, V. Mottl, D. Babichev, I. Pugach, A. Morozov*. Constrained Regularized Regression Model Search in Large Sets of Regressors // Lecture Notes in Artificial Intelligence, Vol. 1035, Springer, 2018, pp. 1-15.

Формирование инвестиционного портфеля по неточной информации о доходностях активов

*Моттль Вадим Вячеславович*¹ vmottl@yandex.ru

*Красоткина Ольга Вячеслав.*² okrasotkina@markovprocesses.com

*Черноусова Елена Олеговна*³ lena-ezhova@rambler.ru

*Рыбка Елизавета Михайловна*⁴ yelizaveta.rybka@phystech.edu

¹Москва, ФИЦ «Информатика и управление» РАН

²Summit, NJ, USA, Markov Processes International

³Москва, Московский физико-технический институт

⁴Париж, Франция, Ecole Polytechnique

DOI: 10.30826/IDP201846

Согласно принципу нобелевского лауреата Гарри Марковица, распределение капитала инвестора по множеству биржевых активов образует эффективный портфель только при соблюдении баланса Парето между двумя противоречивыми желаниями – увеличить капитал и уменьшить риск его потери. Эти два критерия определяются вектором средних значений и ковариационной матрицей случайных будущих доходностей. В итоге оказывается, что эта пара и свободно выбираемый коэффициент склонности инвестора к риску (Risk Tolerance) вместе полностью определяют однопараметрическое семейство портфелей Марковица, которое исчерпывает множество всех «разумных» портфелей. Однако инвестору доступны лишь оценки ожидаемых доходностей и их ковариаций, сами по себе образующие случайную пару «вектор-матрица». В данной работе мы используем распределение Гаусса-Уишарта на множестве таких пар как единое параметрическое семейство априорных и апостериорных распределений, наделяющее понятие эффективного портфеля Марковица способностью учитывать неточность знания будущих доходностей биржевых активов. В частности, мы рассматриваем некоторые специальные случаи при разных сочетаниях степеней недоверия к предполагаемым средним значениям и ковариациям доходностей.

Работа поддержана грантами РФФИ 17-07-00436, 17-07-00993.

- [1] *O. Krasotkina, M. Markov, V. Mottl, D. Babichev, I. Pugach, A. Morozov.* Constrained Regularized Regression Model Search in Large Sets of Regressors // Lecture Notes in Artificial Intelligence, Vol. 1035, Springer, 2018, pp. 1-15.

Investment portfolio composition from uncertain forecast on stock market asset returns

*Vadim Mottl*¹ vmottl@yandex.ru

*Olga Krasotkina*² okrasotkina@markovprocesses.com

*Elena Chernousova*³ lena-ezhova@rambler.ru

*Elizabeth Rybka*⁴ yelizaveta.rybka@phystech.edu

¹Moscow, Computer Science and Control, Russian Academy of Sciences

²Summit, NJ, USA, Markov Processes International

³Moscow Institute of Physics and Technology

⁴Paris, France, Ecole Polytechnique

The commonly adopted principle of investment portfolio composition, formulated by Nobel prize winner Harry Markowitz, states that the capital allocation over a set of stock market assets is effective only under the condition of Pareto balance between two contradictory desires – to enlarge the capital and diminish the risk of losing it. These two criteria are functions of the mean-values vector and covariance matrix of the anticipated random stock market returns. It turns out that such a vector-matrix pair, coupled with a free coefficient of the assumed Risk Tolerance of the investor, jointly define a one-parametric family of Markowitz Portfolios, which completely exhaust the set of all “reasonable” ones. However, the investor can rely only on some estimates of the anticipated returns and their covariances, which, in their turn, make a random vector-matrix pair. In this work, we consider the conjugate family of prior and posterior Gauss-Wishart distributions over the set of such pairs as a means to endow the notion of Markowitz effective portfolio with the ability to perceive the uncertainty of the investor’s knowledge on the asset returns in the next holding period. In particular, we consider some special cases under different combination of reliance degrees put on the expected values of returns and on their covariances.

This research is funded by RFBR, grants 17-07-00436 and 17-07-00993.

- [1] *O. Krasotkina, M. Markov, V. Mottl, D. Babichev, I. Pugach, A. Morozov*. Constrained Regularized Regression Model Search in Large Sets of Regressors // Lecture Notes in Artificial Intelligence, Vol. 1035, Springer, 2018, pp. 1-15.

Алгоритмическая реализация методологии оценивания состава инвестиционных портфелей

*Моттль Вадим Вячеславович*¹ vmottl@yandex.ru

*Красоткина Ольга Вячеслав.*² okrasotkina@markovprocesses.com

*Морозов Алексей Олегович*³ ao.morozov@phystech.edu

*Медведев Алексей Владимирович*⁴ fortunato.mav@gmail.com

¹Москва, ФИЦ "Информатика и управление" РАН

²Summit, NJ, USA, Markov Processes International

³Москва, Московский физико-технический институт

⁴Москва, Московский государственный университет

DOI: 10.30826/IDP201847

С вычислительной точки зрения задача восстановления скрытого состава инвестиционного портфеля по наблюдаемому временному ряду его периодических доходностей есть задача регрессионного анализа при совокупности дополнительных допущений. Во-первых, коэффициенты регрессии (искомое доленое распределение капитала) ограничены неравенствами неотрицательности вместе с равенством единице их суммы. Во-вторых, предполагается, что число регрессоров (число возможных биржевых активов) намного превышает размер выборки. Дополнительное предположение состоит в том, что коэффициенты регрессии отличаются от нуля только в пределах реально существующего малого подмножества регрессоров (искомого состава портфеля) в многократно большем множестве всех активов. В качестве неизбежной проблемно-ориентированной регуляризации мы используем естественное предположение, что анализируемый портфель рационально построен его администрацией в соответствии с теорией эффективных портфелей Гарри Марковица. Мы обеспечиваем линейную вычислительная сложность алгоритма поиска состава портфеля в очень большом множестве всех биржевых активов, в то время, как сложность по относительно небольшому числу наблюдений остается полиномиальной.

Работа поддержана грантами РФФИ 17-07-00436, 17-07-00993.

- [1] *O. Krasotkina, M. Markov, V. Mottl, D. Babichev, I. Pugach, A. Morozov. Constrained Regularized Regression Model Search in Large Sets of Regressors // Lecture Notes in Artificial Intelligence, Vol. 1035, Springer, 2018, pp. 1-15.*

Algorithmic Implementation of Factor Search in Returns Based Analysis of Investment Portfolios

*Vadim Mottl*¹

vmottl@yandex.ru

*Olga Krasotkina*²

okrasotkina@markovprocesses.com

*Alexey Morozov*³

ao.morozov@phystech.edu

*Alexey Medvedev*⁴

fortunato.mav@gmail.com

¹Moscow, Computer Science and Control, Russian Academy of Sciences

²Summit, NJ, USA, Markov Processes International

³Moscow Institute of Physics and Technology

⁴Moscow State University

From the computational point of view, the problem of recovering the hidden sharing of the capital in an investment portfolio from the time series of its periodic returns (Returns Based Style Analysis) is that of regression analysis under some additional assumptions. First, the regression coefficients (the capital shares to be found) are constrained by nonnegativity inequalities and by the unit-sum equality. Second, it is assumed that the number of regressors (number of all stock market assets) far exceeds that of samples. An additional assumption is that the regression coefficients differ from zero only within a really existing small subset (the sought-for portfolio composition) within a much greater universe of all regressors, and the search for this subset (Factor Search) is the main aim of data processing. It is obvious that such a regression problem is ill posed. As a problem-oriented means of the inevitable regularization, we harness the natural assumption that the portfolio is composed by its administration in accordance with the theory of effective portfolios by Harry Markowitz. We provide the linear computational complexity of the Factor Search with respect to the huge number of all stock market assets, whereas the complexity relative the much smaller number of samples remains polynomial.

This research is funded by RFBR, grants 17-07-00436 and 17-07-00993.

- [1] *O. Krasotkina, M. Markov, V. Mottl, D. Babichev, I. Pugach, A. Morozov*. Constrained Regularized Regression Model Search in Large Sets of Regressors // Lecture Notes in Artificial Intelligence, Vol. 1035, Springer, 2018, pp. 1-15.

Семантическая сегментация аэрокосмических изображений промышленных пальмовых плантаций с использованием вполне конволюционных сверточных нейронных сетей

Бакланов А. П.^{1,2}

baklanov@iiasa.ac.at

*Пасынков М. К.*³

pmk@imm.uran.ru

Хачай М. Ю.^{3,4}

mkhachay@imm.uran.ru

¹ Лаксенбург, Австрия, Межд. инст. прикл. сист. анализа

² Санкт-Петербург, Россия, Высшая школа экономики

³ Екатеринбург, Россия, ИММ им. Н.Н.Красовского УрО РАН

⁴ Екатеринбург, Россия, Уральский Федеральный университет

DOI: 10.30826/IDP201848

Исследование мотивировано проблемой мониторинга распространения промышленных пальмовых плантаций (ПП), быстрый рост которых в тропическом поясе Африки, Южной Азии и части Бразилии влечет значительное сокращение ливневых лесов, способствуя глобальному потеплению из-за снижения поглощения углекислого газа. Предлагаемый подход основан на сведении к задаче семантической сегментации аэрофотоснимков земной поверхности, получаемых со спутников Landsat, и применении вполне конволюционных сверточных сетей (FCN). Предлагаемая нами сеть FCN-8s-EL, развивающая архитектуры FCN-8s и VGG-16, значительно опережает традиционные для дистанционного зондирования методы попиксельной классификации, основанные на Random Forest с использованием текстурных признаков, NDVI и всех диапазонов Landsat (ALL, ALL+G(10)). Сеть продемонстрировала высокую устойчивость к пространственным и временным смещениям исходных данных и возможность успешного мониторинга расширения промышленных пальмовых плантаций.

Работа поддержана грантом РФФ № 14-11-00109.

Таблица 1. Сравнение качества алгоритмов (5-fold CV)

| | ALL | ALL+G(10) | FCN-8s-EL |
|--------------------------------|------|-----------|-------------|
| Overall accuracy (OA) | 0.82 | 0.86 | 0.95 |
| OPP precision (UA_1) | 0.71 | 0.73 | 0.95 |
| OPP recall (PA_1) | 0.48 | 0.5 | 0.96 |
| $non-OPP$ precision (UA_0) | 0.87 | 0.89 | 0.99 |
| $non-OPP$ recall (PA_0) | 0.95 | 0.95 | 0.98 |

- [1] *Baklanov A. et al.* Application of Fully Convolutional Neural Networks to Mapping Industrial Oil Palm Plantations // LNCS, 2018, Vol. 11179.

Semantic segmentation of industrial oil palm plantations using Fully Convolutional Neural Networks

Baklanov A.^{1,2}

baklanov@iiasa.ac.at

*Pasynkov M.*³

pmk@imm.uran.ru

Khachay M.^{3,4}

mkhachay@imm.uran.ru

¹ Laxenburg, Austria, Internatioanl Inst. for Applied Systems Analysis

² St. Petersburg, Russia, Higher School of Economics

³ Ekaterinburg, Russia, Krasovsky Inst. of Mathematics and Mechanics

⁴ Ekaterinburg, Russia, Ural Federal University

This research is motivated by sustainability problems of oil palm expansion. Fast-growing industrial Oil Palm Plantations (OPPs) in the tropical belt of Africa, Southeast Asia and parts of Brazil lead to significant loss of rainforest and contribute to the global warming by the corresponding decrease of carbon dioxide absorption. We propose a novel approach to monitoring of the expansion of OPPs based on an application of state-of-the-art Fully Convolutional Neural Networks (FCNs) to solve Semantic Segmentation Problem for Landsat imagery. The proposed approach (FCN-8s-EL) significantly outperforms per-pixel classification methods based on Random Forest using texture features, NDVI, and all Landsat bands (ALL, ALL+G(10)). Moreover, the trained FCN is robust to spatial and temporal shifts of input data. The paper provides a proof of concept that FCNs as semi-automated methods enable OPPs mapping of entire countries and may serve for yearly detection of oil palm expansion.

This research was supported by Russian Science Foundation, grant no. 14-11-00109.

Table 1. Results obtained by 5-fold cross-validation

| | ALL | ALL+G(10) | FCN-8s-EL |
|------------------------------|------|-----------|-------------|
| Overall accuracy (OA) | 0.82 | 0.86 | 0.95 |
| OPP precision (UA_1) | 0.71 | 0.73 | 0.95 |
| OPP recall (PA_1) | 0.48 | 0.5 | 0.96 |
| non-OPP precision (UA_0) | 0.87 | 0.89 | 0.99 |
| non-OPP recall (PA_0) | 0.95 | 0.95 | 0.98 |

- [1] *Baklanov A. et al.* Application of Fully Convolutional Neural Networks to Mapping Industrial Oil Palm Plantations // Analysis of Images, Social Networks and Texts: Revised Selected Papers 7th Internat. Conf. (AIST-2018), Cham: Springer International Publishing, 2018, LNCS, Vol. 11179.

Метод проекций в пространстве лучевого преобразования Радона для детектирования пешеходов

Гнеушев Александр Николаевич^{1*} gneushev@ccas.ru

*Самсонов Никита Андреевич*² nikita.samsonov@phystech.edu

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

²Москва, Московский физико-технический институт (ГУ)

DOI: 10.30826/IDP201849

Задача автоматического обнаружения человека и других объектов на изображениях востребована в таких областях как видео аналитика, видеонаблюдение и дорожный контроль, индексирование изображений в больших базах данных.

В работе развивается подход построения локального дескриптора для детектирования объектов на основе лучевого преобразования Радона градиентного поля изображения. Пространство лучевого преобразования Радона рассматривается как аккумуляторное пространство Хафа, в котором строятся проекции (НАН, Nough Accumulator Histograms). Совокупность локальных проекций формирует НАН-дескриптор области изображения объекта, обобщающий НОГ-дескриптор (Histograms of Oriented Gradients).

В работе показывается, что вычисление НАН-дескриптора может быть произведено по схеме сверточной нейронной сети специального вида. Исследуются вопросы влияния точности аппроксимации функции вклада в преобразовании Радона, вида локальной нормализации, количества бинов в проекционных гистограммах на результат детектирования пешеходов. Так же оценивается устойчивость НАН-детектора от вариаций масштаба детектируемого объекта. Приведены результаты работы НАН-дескриптора в сравнении с НОГ-дескриптором, а так же с результатами работы сверточной нейронной сети архитектуры ResNeXt на базах INRIA и Cityscapes.

Работа поддержана грантом РФФИ № 16-07-01171.

- [1] *Гнеушев А. Н., Самсонов Н. А.* Метод проекций в пространстве лучевого преобразования Радона градиентного поля изображения для детектирования пешеходов // Машинное обучение и анализ данных, 2018.

The method of projections in the Radon ray transform space for pedestrian detection

Gneushev Alexander^{1*}

gneushev@ccas.ru

*Samsonov Nikita*²

nikita.samsonov@phystech.edu

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

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

The task of automatic pedestrian and other objects detection on images is actual in such areas as video analytics, video surveillance and road monitoring, image indexing in large databases.

In the article the approach of constructing a local descriptor for object detection based on the Radon ray transform of the gradient image field is expanded. The Radon ray transform space is considered as the Hough accumulator space, in which the projections (HAH, Hough Accumulator Histograms) are constructed. The set of local projections defines the HAH-descriptor of the object image area, which generalizes the HOG-descriptor (Histograms of Oriented Gradients). In the article it is shown that the calculation of the HAH-descriptor can be performed according to a Convolutional Neural Network approach with a special architecture. The questions of the influence of the Radon transform contribution function approximation accuracy, the type of local normalization, the number of the projection histograms bins on the pedestrians detection result are investigated. The stability of the HAH-detector depended on the object scale variations in the image is also estimated. The results of the HAH-descriptor in comparison with the HOG-descriptor and the ResNeXt Convolution Neural Network on INRIA and Cityscapes databases are presented.

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

- [1] *Gneushev A., Samsonov N.* The method of projections in the Radon ray transform space of the gradient image field for pedestrian detection // Journal of Machine Learning and Data Analysis, 2018.

Алгоритм сжатия динамического диапазона HDR изображений на основе фильтрации с сохранением структуры

Грачева Инесса Александровна^{1*}

gia1509@mail.ru

Копылов Андрей Валериевич¹

And.Kopylov@gmail.com

¹Тула, Тульский государственный университет

DOI: 10.30826/IDP201850

В настоящее время все большее распространение получают камеры высокого разрешения с технологией съемки изображений с большим динамическим диапазоном (High dynamic range, HDR). Однако, большинство современных устройств отображения не могут полностью отобразить все оттенки HDR изображений (до 100000:1), поскольку их динамические диапазоны намного меньше (около 1000:1). В связи с ростом пространственного разрешения камер задача разработки быстродействующего метода отображения HDR изображений без потери качества остается актуальной. В данной работе предлагается метод тонового отображения HDR изображений с использованием фильтрации с сохранением структуры на основе вероятностной гамма-нормальной модели [1]. В рамках предлагаемого подхода значения интенсивности извлекаются из исходного HDR изображения в логарифмической шкале. Базовый слой изображения вычисляется из логарифмического с помощью фильтрации с сохранением структуры исходного HDR изображения на основе вероятностной гамма-нормальной модели, а детальный слой получается вычитанием базового слоя из логарифмического. Затем специальный оператор отображения тонов преобразует базовый и детальный слои в карту тонов изображения, на основе которой осуществляется сжатие динамического диапазона с сохранением деталей. Предлагаемый метод имеет сопоставимое качество и наименьшее время вычислений по сравнению с другими операторами отображения тонов. Работа поддержана грантом РФФИ № 16-07-01039, 16-57-52042.

- [1] Грачева И. А., Копылов А. В. Алгоритм сжатия динамического диапазона HDR изображений на основе фильтрации с сохранением структуры // Изв. ТулГУ, Техн. науки Тула: Изд-во ТулГУ, 2018. (в печати)

Dynamic range compression algorithm of HDR images on the basis of the structure-preserving filtering

Gracheva Inessa^{1*}

gia1509@mail.ru

*Kopylov Andrey*¹

And.Kopylov@gmail.com

¹Tula, Tula State University

High-resolution cameras with high dynamic range (HDR) image capture technology are now becoming increasingly popular. However, most modern display devices cannot fully display all the shades of HDR images (up to 100000:1) because their dynamic ranges are much smaller (about 1000:1). Taking into account the growth of spatial resolution of cameras, the task of developing a high-speed method of HDR imaging remains relevant.

In this paper, we propose the tone mapping method for HDR images with using structure-preserving filtering on the basis of the probabilistic gamma-normal model [1]. In the framework of the proposed approach, intensity values are extracted from the original HDR image in a logarithmic scale. A base layer of the image is calculated from the logarithmic layer by the structure-preserving filtering of the original HDR image on the basis of the probabilistic gamma-normal model, and a detail layer is obtained by subtracting the base layer from the logarithmic one. Then, the special tone mapping operator converts base and detail layers into a tone map of the image, on which the dynamic range compression with preserving details is done. The proposed method has comparable quality and least computation time in comparison to other tone mapping operators.

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

- [1] *Gracheva I., Kopylov A.* Dynamic range compression algorithm of HDR images on the basis of the structure-preserving filtering // TSU News, Technical science, Tula: Publisher of TSU, 2018. (in press).

Модель скоростной кластеризации пикселей изображения на основе метода Уорда

Ханыков Игорь Георгиевич¹*

igk@iiias.spb.su

¹Санкт-Петербург, Институт информатики и автоматизации
Российской академии наук

DOI: 10.30826/IDP201851

Среди применимых в обработке изображений методов кластерного анализа особое место занимает классический метод Уорда, который обрабатывает цветные изображения и возвращает адекватные результаты. Но характерная методу высокая вычислительная сложность существенно ограничивает его применение. Рассматриваемая в работе [1] модель преодолевает указанный недостаток за счет разделения процесса обработки на три типовых этапа. Первый этап быстро строит грубую иерархию связанных сегментов, генерация которой доступна двумя разными вариантами. Первый вариант – применение модели Мамфорда-Шаха для укрупнения сегментов на каждом шаге. Второй вариант – разделение изображения на фрагменты регулярной сеткой для их обработки как самостоятельные изображения классическим методом Уорда с последующим объединением иерархий в одну. Вторым этапом формирует заданное число суперпикселей, фактически, выполняя улучшение качества заданного разбиения при фиксированном числе цветов (кластеров). Для чего разработаны два базовых алгоритма: SI-метод (Segmentation Improvement) и K-meanless метод (метод K-средних-без-средних). Множество программных реализаций обусловлено как возможностью комбинированного сочетания пары базовых методов SI и K-meanless (отдельно, последовательно, циклично), так и версиями самих методов (сегментарная, кластерная). Третий этап выполняет кластеризацию суперпикселей методом Уорда. Экспериментальные результаты, описание этапов и операций представлены в докладе.

- [1] *Khanykov I. G., Kharinov M. V., Patel C.* Image segmentation improvement by reversible segment merging // *Soft Computing and its Engineering Applications (icSoftComp), International Conference on* – IEEE, 2017. — С. 1–8. DOI: 10.1109/ICSFTCOMP.2017.8280096

The Model of the High-Speed Clustering of the Image Pixels on Ward's Method Basis

*Khanykov Igor*¹★

igk@iiias.spb.su

¹Saint Petersburg, Institute for Informatics and Automation of the Russian Academy of Sciences

Among the methods of cluster analysis that are applicable in image processing the classical Ward's method, which processes the color images and returns adequate results, takes a special place. But the high computational complexity characteristic of the method considerably limits its use. The considered in [1] model of the high-speed clustering of the image pixels overcomes the indicated drawback by virtue of splitting the image processing into three typical stages.

The first stage quickly builds a coarse hierarchy of adjacent segments, which generation is available in two different ways. The first option is to use the Mumford-Shah model to enlarge the segments at each step. The second option is the division of the image into fragments by a regular grid for their processing as stand-alone images using the classical Ward's method followed by merging of hierarchies into one.

The second stage forms a given number of superpixels, in fact, performing a quality improvement of the specified partition for a fixed number of colors (clusters). Two basic algorithms have been developed for this: SI-method (Segmentation Improvement) and K-meanless method (K-means-without-average). The variety of software realizations is specified both by the combinational possibilities of the pair of basic SI and K-meanless methods (separately, sequentially, cyclically), and by the versions of the methods themselves (segmental, clusteral).

The third stage is the clustering of the superpixels using the Ward's method.

The experimental results, description of the stages and the operations are presented in the report.

- [1] *Khanykov I. G., Kharinov M. V., Patel C.* Image segmentation improvement by reversible segment merging // *Soft Computing and its Engineering Applications (icSoftComp)*, International Conference on. – IEEE, 2017. — pp. 1–8. DOI: 10.1109/ICSOFTECOMP.2017.8280096

Об использовании инвариантов отражения в практическом решении задачи определения зеркальной симметрии

Федотова Софья Антоновна `fedotova.sonya@gmail.com`

Середин Олег Сергеевич `oseredin@yandex.ru`

Кушнир Олеся Александровна `kushnir-olesya@rambler.ru`

Россия, Тула, Тульский государственный университет

DOI: 10.30826/IDP201852

Поиск осей зеркальной симметрии на бинарных изображениях является вычислительно сложной задачей. В исследовании [1] предлагается использовать инварианты отражения для определения самого факта наличия зеркальной симметрии, что позволяет выявлять асимметричные объекты на изображениях еще до запуска процедуры поиска оси симметрии. В работе [2] экспериментально проверяется применимость инвариантов для определения зеркальной симметрии на реальных изображениях. Для этого оценки симметричности, полученные с использованием инвариантов отражения, сравниваются с объективной мерой, рассчитанной на основе подобия Жаккара. Экспериментальные исследования, проведенные на базах бинарных растровых изображений Бабочки и MPEG 7 CE Shape-1 Part B, показали невозможность использования инвариантов отражения для быстрого выявления приближенной симметрии на бинарных растровых изображениях.

Работа выполнена при поддержке РФФИ, гранты № 16-57-52042, № 18-07-00942.

- [1] *Li E., Li H.* Reflection Invariant and Symmetry Detection // arXiv preprint arXiv:1705.10768, 2017.
- [2] *Федотова С. А., Середин О. С., Кушнир О. А.* Об использовании инвариантов отражения в практическом решении задачи определения зеркальной симметрии // Известия Тульского государственного университета. Технические науки., Тула: Издательство ТулГУ, 2018 (в печати).

On practical solution to the problem of reflection symmetry detection using reflection invariants

*Fedotova Sofia**

fedotova.sonya@gmail.com

Seredin Oleg

oseredin@yandex.ru

Kushnir Olesia

kushnir-olesya@rambler.ru

Tula, Russia, Tula State University

The task of reflection symmetry axis searching applied to binarized real-world images is very time-consuming. In [1], it is proposed to use reflection invariants for symmetry detection. This would allow us to determine asymmetric objects in images before starting the procedure of symmetry axis searching. In [2], the possibility of applying reflection invariants to the task of symmetry detection on binarized real-world images is experimentally tested. To do this, we compare values of reflection invariants to the objective symmetry measure calculated on the basis of Jaccard similarity. Experimental study was performed on Butterfly and MPEG 7 CE Shape-1 Part B datasets. Results show that it is impossible to use reflection invariants for fast detection of approximate symmetry on binary raster images.

This research is funded by RFBR, grants 16-57-52042, 18-07-00942.

- [1] *Li E., Li H.* Reflection Invariant and Symmetry Detection // arXiv preprint arXiv:1705.10768, 2017.
- [2] *Fedotova S., Seredin O., Kushnir O.* On practical solution to the problem of reflection symmetry detection using reflection invariants // *Izvestiya Tula State University. Technical science, Tula, 2018 (in press).*

Задача распознавания символов в художественных изображениях

Лемтюжникова Дарья Владимировна darabbt@gmail.com

Москва, МФТИ

DOI: 10.30826/IDP201853

Генерация потока новых произведений живописи и их электронных изображений в Интернете приводит к необходимости автоматизации их первичного искусствоведческого анализа, в частности, описания художественных произведений в терминах символов. Понятие «символа» в искусствоведении является одним из ключевых и, в то же время, одним из самых трудно формализуемых. Наиболее естественным является прецедентный способ его формализации — экспертная разметка изображений и формирование обучающей выборки для решения задачи классификации (тегирования) изображений. В качестве тегов предлагается использовать библиотеку символов <http://www.symbolarium.ru> на русском языке и <https://www.symbols.com> на английском. Кроме того, предлагается использовать текстовые описания как самих произведений, так и статей энциклопедии символов.

Для решения задачи тегирования изображений предлагается использовать глубокие свёрточные нейронные сети, мульти-модальные и мультиязычные вероятностные тематические модели, а также методы решения разреженных задач дискретной оптимизации. Задача тегирования заключается в том, чтобы построить отношения многие-ко-многим между тремя конечными множествами: множеством символов-тегов S , множеством ключевых слов W и множеством выходных нейронов сети H , интерпретируемых как детекторы ключевых элементов изображений. Предлагается использовать локальный элиминационный алгоритма (ЛЭА), предложенный О. А. Щербиной. ЛЭА сводит исходную задачу большой размерности к последовательности задач меньших размерностей. В [1] проведены эксперименты по исследованию вычислительной эффективности ЛЭА.

- [1] *Лемтюжникова Д. В., Ковков Д. В.* Тестирование алгоритмов для целочисленных квазиблочных задач оптимизации // Вестник МГТУ им. Н.Э.Баумана, серия Информационные технологии. 2017. №6.

Towards recognition of symbols in painting images

Darya Lemtyuzhnikova

darabbt@gmail.com

Moscow, MIPT

The enormous flow of new paintings and their electronic images in the Internet evoke the need of automatization of their primary analysis from the art history point of view, particularly in terms of symbols. The key concept of “symbol” is one of the keystone in the art and one of the most difficult to formalize at the same time. The precedent-based approach seems to be most natural for the formalization. It implies the creation of a training set of images tagged by symbols and considering the tagging problem as a machine learning task. We propose to use the libraries of symbols, signs, and glyphs, arranged by categories such as culture, country, religion, and more, such as <http://www.symbolarium.ru> in Russian and <https://www.symbols.com> in English. We propose to use text descriptions of paintings themselves and the articles from the encyclopedia of symbols in order to better associate images and symbols.

We propose to use deep convolutional neural networks, multimodal and multilingual probabilistic topic models, and sparse discrete optimization to tackle the problem of tagging images by symbols. The tagging problem is to build a sparse many-to-many relationships between three finite sets: a set S of symbol tags, a set W of keywords, and a set H of output neurons of the convolution network interpreted as graphical primitive detectors. We proposed to use a local elimination algorithm (LEA), proposed by O.A. Shcherbina. The LEA reduces the original large-dimensional problem to a sequence of problems of smaller dimensions. We show the computational efficiency of LEA experimentally in [1].

- [1] *Lemtyuzhnikova D.* Parallel representation of local elimination algorithm for accelerating the solving sparse discrete optimization problems // Computer Research and Modeling, Moscow: Institute of computer science, 2015. — vol. 7, no. 3. — p.699–705 (in Russian).

Измерение плотности холста картин по изображениям

Мурашов Дмитрий Михайлович^{1*}

d_murashov@mail.ru

*Березин Алексей Владимирович*²

berezin_aleks@mail.ru

*Иванова Екатерина Юрьевна*³

ivanova-e-u@yandex.ru

¹Москва, ФИЦ ИУ РАН

²Москва, ГИМ

³Москва, РАЖВиЗ Ильи Глазунова

DOI: 10.30826/IDP201854

Работа посвящена решению задачи подсчета количества нитей на изображениях холстов картин, что необходимо для определения характеристик, используемых искусствоведами для датировки произведений живописи.

В последние несколько лет были разработаны автоматизированные алгоритмы вычисления характеристик холстов по рентгеновским и высококачественным терагерцовым изображениям. Для контроля плотности ткани в текстильном производстве используются фотографии, полученные при просвечивании участка ткани источником света. Особенностью нашего исследования является использование изображений, полученных фотосъемкой при направленном под острым углом по отношению к холсту освещении. Такой способ получения изображений позволил подчеркнуть текстуру холста в выбранном направлении.

Для анализа изображений образцов предложены модификации известного алгоритма, основанного на фильтрации в Фурье-области и пороговой бинаризации, а также новый алгоритм на основе локализации хребтов полутоновых изображений. В известных работах количество нитей определяется по пикам спектра Фурье или по базовым линиям на изображении холста. В предлагаемой работе подсчет нитей выполняется по всем строкам/столбцам матрицы изображения, и по результатам строится гистограмма. Искомое число нитей определяется по максимуму полученной гистограммы. Применение гистограмм позволяет снизить влияние на результат артефактов, полученных при обработке изображений. Для пороговой бинаризации применялись методы Отсу и Ниблака.

Measuring canvas density from images

*Murashov Dmitry*¹★

d_murashov@mail.ru

*Berezin Aleksey*²

berezin_aleks@mail.ru

*Ivanova Ekaterina*³

ivanova-e-u@yandex.ru

¹Moscow, FRC CSC RAS

²Moscow, State Historical Museum

³Moscow, Glazunov Academy

This paper deals with the problem of painting thread counting from images. It is necessary to determine the characteristics used by art historians for dating works of art.

In the last few years, automated algorithms for calculating canvas characteristics from x-ray and high-quality terahertz images have been developed. To control the fabric density in textile industry, microscopic photographs, obtained when the fabric sample is illuminated by a light transmitted source, are used. The peculiarity of our research is acquiring canvas images in raking light. This way of acquiring images allowed to emphasize the texture of the canvas in the specified direction. For the analysis of canvas sample images we propose modifications of known algorithm based on a filtering in the Fourier domain and thresholding, and the new algorithm based on localizing grayscale image ridges. In known works, the number of threads is determined by the Fourier spectrum peaks or by the baselines in the canvas image. In this paper, the counting of threads is performed over all rows / columns of the image matrix, and a histogram is constructed based on the results. The desired number of threads is determined by the maximum of the histogram obtained. The use of histograms allows to reduce inaccuracy produced by artifacts obtained during image processing. For thresholding, Otsu and Niblack methods are applied.

A computing experiment on the study of canvases of five portraits of the work of Russian artists of the 18th century was carried out. The results of the experiment show the following. The algorithm based on the Otsu method does not require parameters and has acceptable accuracy and high speed.

Проведен вычислительный эксперимент по исследованию холстов пяти портретов работы русских художников XVIII века. Результаты проведенных исследований позволили сделать следующие выводы. Алгоритм на основе метода Отсу не имеет параметров и обладает хорошей точностью и высоким быстродействием. Однако на нескольких изображениях этот алгоритм дал неприемлемый результат. Алгоритм на базе метода Ниблака требует настройки двух параметров и больших вычислительных затрат по сравнению с алгоритмом с глобальным методом бинаризации, но показал в среднем более высокую точность определения плотности. Алгоритм измерения на основе хребтов полутонового изображения требует настройки большего количества параметров и существенно больших вычислительных затрат, по сравнению с другими алгоритмами, но показал наилучший результат по точности измерения в пределах допуска, приемлемого для экспертизы и атрибуции. Исследованные алгоритмы обеспечивают точность измерения плотности холста в пределах одной нити на сантиметр на 70 – 97 процентах изображений образцов.

Результаты вычислительного эксперимента соответствуют результатам известных алгоритмов измерения плотности холста по рентгеновским изображениям картин. Для повышения надежности измерений плотности холста при исследованиях картин предпочтительно использовать несколько алгоритмов.

Дальнейшие исследования будут направлены на повышение быстродействия алгоритмов и повышение точности.

Полная версия работы представлена в [1].

Работа поддержана грантом РФФИ № 18-07-01385.

- [1] *Мурашов Д. М., Березин А. В., Иванова Е. Ю.* Определение количества нитей холстов картин по изображениям, полученным при направленном освещении // Машинное обучение и анализ данных, Москва: ФИЦ ИУ РАН, 2018.

However, on several images this algorithm gave an unacceptable result. The algorithm based on the Niblack method requires setting up two parameters and computationally expensive compared to the algorithm with the global threshold method, but showed on average a higher density measurement accuracy. The measurement algorithm based on localizing grayscale image ridges requires setting more parameters and significantly higher computational costs than other algorithms, but has shown the best result in measuring accuracy within the error limit acceptable for expertise and attribution of paintings. The researched algorithms provide the accuracy of measuring the canvas density from within one thread per centimeter for 70 – 97 percents of the sample images. The results of the computing experiment correspond to the results of known algorithms for measuring canvas density from x-ray images of paintings. To improve the reliability the canvas density measurements in painting analysis is preferable to use several algorithms.

Further research will be aimed at improving the accuracy and speed of the algorithms.

The full version of this work is presented in paper [1]

This research is funded by RFBR, grant 18-07-01385.

- [1] *Murashov D., Berezin A., Ivanova E.* Painting canvas thread counting from images obtained in raking light // Machine Learning and Data Analysis, M: FRC CSC RAS, 2018.

Измерение пространственных спектров морских волн по оптическим аэрокосмическим изображениям высокого разрешения

Мурынин Александр Борисович^{1,2}

AMurynin@bk.ru*

*Бондур Валерий Григорьевич*¹

VGBondur@aerocosmos.info

¹Москва, НИИ АЭРОКОСМОС

²Москва, ФИЦ ИУ РАН

DOI: 10.30826/IDP201855

Работа посвящена методам дистанционного измерения пространственных спектров волн, возникающих на поверхности морей и океанов. Пространственные спектры, получаемые при обработке аэрокосмических изображений, позволяют исследовать различные физические процессы и явления в системе океан-атмосфера на обширных акваториях в широком диапазоне пространственных частот. Проведён анализ физических механизмов, ответственных за формирование оптических аэрокосмических изображений взволнованной морской поверхности. Показано, что в большинстве естественных условий формирования таких изображений происходит нелинейная модуляция поля яркости уклонами элементов водной поверхности. Предложены методы восстановления спектров поверхностного волнения по спектрам оптических изображений с учётом такой модуляции, основанные на численном моделировании изображений взволнованной водной поверхности с учётом условий волнообразования и условий поступления света к морской поверхности. С использованием результатов такого моделирования строятся специальные операторы для восстановления спектров волнения по спектрам аэрокосмических изображений, зависящие от наборов параметров, определяемых условиями формирования изображений. Приведены результаты экспериментальных исследований спектров морских волн в различных акваториях с использованием космических оптических изображений высокого пространственного разрешения. Работа поддержана Минобрнауки РФ, проект № 5.9330.2017/8.9.

- [1] Bondur V.G., Murynin A.B. Measurement of Sea Wave Spatial Spectra from High-Resolution Optical Aerospace Imagery // in book Surface Waves, IntechOpen, 2018. — С. 71–78. <http://www.intechopen.com/books/surface-waves-new-trends-and-developments>

Measurement of sea wave spatial spectra from high-resolution optical aerospace imagery

Alexander Murynin^{1,2*}

AMurynin@bk.ru

*Valery Bondur*²

VGBondur@aerocosmos.info

¹Moscow, ISR AEROCOSMOS

²Moscow, FRC CSC RAS

The work is devoted to the development of methods for remote measurement of spatial spectra of waves arising on the surface of seas and oceans. The spatial spectra obtained during the processing of aerospace images allow us to study various physical processes and phenomena in the ocean-atmosphere system over vast areas of water in a wide range of spatial frequencies. The analysis of the physical mechanisms responsible for the formation of optical images of the rough sea surface recorded by remote sensing instruments is carried out. It is shown that in most natural conditions of such image formation, a nonlinear modulation of the brightness field occurs by slopes of water surface elements. Methods for reconstructing the spectra of surface waves from optical image spectra with allowance for such modulation are proposed. The methods are based on the numerical modeling of images of an agitated water surface taking into account wave formation conditions and conditions of light entering the sea surface from the upper and lower hemispheres. The results of numerical modeling of water surface images under various external conditions are presented. Using the results of this simulation, special operators are built to retrieve wave spectra from the spectra of aerospace images. These retrieving operators are presented in the form of analytical expressions, depending on the sets of parameters, which are determined by the conditions for the formation of images. The results of experimental studies of the sea wave spectra in various water areas using satellite optical images of high spatial resolution are presented. This research is supported by Ministry of Education and Science of the Russian Federation, project 5.9330.2017/8.9.

- [1] *Bondur V.G., Murynin A.B.* Measurement of Sea Wave Spatial Spectra from High- Resolution Optical Aerospace Imagery // in book *Surface Waves*, IntechOpen, 2018. — p. 71–78. <http://www.intechopen.com/books/surface-waves-new-trends-and-developments>

Поиск оптимальных параметров вероятностного алгоритма повышения пространственного разрешения мультиспектральных спутниковых изображений

Мурынин Александр Борисович^{1,2*}

AMurynin@bk.ru

*Гороховский Константин Юрьевич*¹

Игнатьев Владимир Юрьевич^{1,2}

vladimir.ignatiev.mipt@gmail.com

*Ракова Кристина Олеговна*³

¹Москва, НИИ АЭРОКОСМОС

²Москва, ФИЦ ИУ РАН

³Москва, МФТИ

DOI: 10.30826/IDP201856

Предложен вероятностный метод повышения пространственного разрешения мультиспектральных космических изображений с использованием опорного изображения. Вычисляется математическое ожидание яркости пикселя в различных каналах улучшенного мультиспектрального изображения на основе вероятностных характеристик окрестности пикселя на мультиспектральном изображении и общей интенсивности яркости панхроматического изображения в этой точке. Проводится анализ применимости различных метрик оценки качества улучшения пространственного разрешения спутниковых изображений. Используется набор наиболее адекватных метрик оценки качества, строится оптимизационная процедура для настройки параметров предложенного вероятностного метода повышения разрешения. Представлены результаты тестирования метода на мультиспектральных изображениях, полученных с различных спутников в различных пространственных разрешениях. Работоспособность алгоритма проверена на различных масштабах увеличения. Проведен сравнительный анализ результатов работы предложенного метода с аналогичными подходами. Работа поддержана РФФИ, грант № 16-51-55019.

- [1] *К. Ю. Гороховский, В. Ю. Игнатьев, А. Б. Мурынин, К. О. Ракова* Поиск оптимальных параметров вероятностного алгоритма повышения пространственного разрешения мультиспектральных спутниковых изображений // Известия РАН. Теория и системы управления, 2017, № 6. — С. 112–124.

Parameters Optimization of the Novel Probabilistic Algorithm for Improving Spatial Resolution of Multispectral Satellite Images

Alexander Murynin^{1,2*}

AMurynin@bk.ru

*Konstantin Gorokhovskiy*¹

Vladimir Ignatiev^{1,2}

vladimir.ignatiev.mipt@gmail.com

*Kristina Rakova*³

¹Moscow, ISR AEROCOSMOS

²Moscow, FRC CSC RAS ³Moscow, MIPT

A probabilistic method for improving the spatial resolution of multispectral space images using a reference image is proposed. The developed method calculates the mathematical expectation of pixel brightness in different channels of an improved multispectral image based on the probabilistic characteristics of the pixel neighborhood on the multispectral image and the overall brightness intensity of the panchromatic image at that point. The applicability of different metrics for evaluating the quality of the spatial resolution of satellite images is analyzed. A set of the most adequate quality evaluation metrics is used. An optimization procedure is developed to adjust the parameters of the proposed probabilistic resolution improvement method. The results of testing the method on the multispectral images obtained from different satellites in different spatial resolutions are presented. The efficiency of the algorithm is tested at different magnification scales. A comparative analysis of the results of the proposed method with similar approaches is conducted. This research is supported by RFBR, grant 16-51-55019.

- [1] *Bondur V.G., Murynin A.B.* Parameters Optimization of the Novel Probabilistic Algorithm for Improving Spatial Resolution of Multispectral Satellite Images // Journal of Computer and Systems Sciences International, 2017, Vol. 56, No. 6, — p. 1008–1020.

Разработка и исследование дважды стохастических моделей случайных полей для представления и обработки изображений со сложной структурой

Андрьянов Никита Андреевич^{1,2*} nikita-and-nov@mail.ru

*Васильев Константин Константинович*¹ vkk@ulstu.ru

¹Ульяновск, Россия, Ульяновский государственный технический университет

²Ульяновск, Россия, Ульяновский институт гражданской авиации имени Главного маршала авиации Б. П. Бугаева

DOI: 10.30826/IDP201857

Работа посвящена синтезу и анализу нового типа моделей изображений, позволяющего адекватно описывать неоднородный в пространстве реальный материал. Для улучшения качества описания изображений вводятся динамические параметры модели. Такой подход позволяет адаптироваться к изменениям, встречающимся на реальных изображениях. При синтезе используется математический аппарат для авторегрессионных случайных полей, что позволяет получить достаточно простое описание модели. Предложена методика формирования изображений с изменяющимися свойствами на основе оценивания параметров в скользящем окне. Приведен ряд примеров генерируемых на основе дважды стохастической модели изображений. Особое внимание уделено задачам оценивания параметров моделей. Описана методика, позволяющая по реальным изображениям формировать дважды стохастическую модель. Также рассмотрены алгоритмы обнаружения точечных аномалий на фоне дважды стохастических случайных полей. Разработан алгоритм обнаружения детерминированных аномалий, обеспечивающий выигрыш по сравнению с обнаружителем на базе авторегрессионных моделей [1]. Работа поддержана грантами РФФИ №№ 17-01-00179 и 18-31-00056.

- [1] *Vasiliev K. K., Andriyanov N. A.* Synthesis and analysis of doubly stochastic models of images // 2nd International Workshop on Radio Electronics and Information Technologies, REIT 2017, Yekaterinburg; Russian Federation; CEUR Workshop Proceedings, Volume 2005, 2017, p. 145-154 <http://ceur-ws.org/Vol-2005/paper-17.pdf>.

Development and investigation of doubly stochastic random fields for representation and processing images with complex structure

Andriyanov Nikita^{1,2*}

nikita-and-nov@mail.ru

*Vasiliev Konstantin*¹

vkk@ulstu.ru

¹Russia, Ulyanovsk, Ulyanovsk State Technical University

²Russia, Ulyanovsk, Ulyanovsk Institute of Civil Aviation

The research is devoted to the synthesis and analysis of a new type of image models, which makes it possible to adequately describe the real material, which is inhomogeneous in space. To improve the quality of image description, it is proposed to introduce the dynamic parameters of the model. This approach allows you to adapt to changes that occur on real images. The synthesis of such models requires a mathematical apparatus for autoregressive random fields, which makes it possible to obtain a quite simple description of the model. The universal technique for creating images with varying correlation properties based on estimating parameters in a sliding window has been proposed. We give a number of examples of images generated by using doubly stochastic model. Particular attention is paid to the problems of estimating the models' parameters. A technique is described that allows for real images to form a doubly stochastic model and its implementation. An algorithm for identifying parameters by real data is proposed. This algorithm is based on a two-dimensional autoregressive model with multiple roots of the characteristic equations of multiplicity (2.2). Algorithms for detecting point anomalies against a background of doubly stochastic random fields are also considered. An algorithm for detecting deterministic anomalies has been developed. The algorithm provides a gain in comparison with the detector based on autoregressive models [1].

This research is funded by the Russian Foundation for Basic Research, grants 17-01-00179 and 18-31-00056.

- [1] *Vasiliev K. K., Andriyanov N. A.* Synthesis and analysis of doubly stochastic models of images // 2nd International Workshop on Radio Electronics and Information Technologies, REIT 2017, Yekaterinburg; Russian Federation; CEUR Workshop Proceedings, Volume 2005, 2017, p. 145-154 <http://ceur-ws.org/Vol-2005/paper-17.pdf>.

Метод детектирования подделок в мобильной системе распознавания по радужной оболочке

Ефимов Юрий Сергеевич^{1*} yuri.efimov@phystech.edu

*Одиноких Глеб Андреевич*² g.odinokikh@gmail.com

*Соломатин Иван Андреевич*¹ ivan.solomatin@phystech.edu

*Коробкин Михаил Владимирович*³
mikhail.korobkin@hotmail.com

*Матвеев Иван Алексеевич*² matveev@ccas.ru

¹Москва, Московский Физико-Технический Институт

²Москва, ФИЦ ИУ РАН

³Зеленоград, Национальный исследовательский университет "МИЭТ"

DOI: 10.30826/IDP201858

Для определения живости глаза на представленном изображении в данной работе предлагается метод, основанный на применении небольшой свёрточной нейронной сети, способный обрабатывать входные данные на мобильном устройстве с ограниченными вычислительными возможностями в реальном времени. Классификация осуществляется с использованием пары растров, соответствующим области глаза на исходном изображении и нормализованной радужке. Рассматриваются следующие типы подделывания текстуры радужных оболочек: распечатки инфракрасных снимков на бумаге, распечатки, покрытые прозрачной контактной линзой, и распечатки, покрытые прозрачным клеем. Оценка качества работы предлагаемого метода осуществляется при помощи собранной вручную базы изображений. Метод показывает высокую точность классификации и скорость обработки входных данных, а также устойчиво работает в переменных условиях освещённости.

Работа поддержана грантом РФФИ № 16-07-01171.

- [1] *Ефимов Ю. С.* Метод детектирования подделок в мобильной системе распознавания по радужной оболочке // Pattern Recognition and Image Analysis, Moscow: Pleiades Publishing, Ltd., 2018 (в печати).

Iris Anti-spoofing Solution for Mobile Biometric Applications

*Efimov Yuri*¹ *

yuri.efimov@phystech.edu

*Odinokikh Gleb*²

g.odinokikh@gmail.com

*Solomatin Ivan*¹

ivan.solomatin@phystech.edu

*Korobkin Mikhail*³

mikhail.korobkin@hotmail.com

*Matveev Ivan*²

matveev@ccas.ru

¹Moscow, Moscow Institute of Physics and Technology

²Moscow, FRC CSC RAS

³Zelenograd, National Research University of Electronic Technology

The ability to provide reliable protection against counterfeiting is one of the key requirements for a biometric security system. Iris recognition as the technology emerging on mobile market is assumed to handle various types of spoof attacks to prevent compromise of the user's personal data.

The method of iris anti-spoofing is proposed in this work. It's based on applying of a convolutional neural network and capable to work in real-time on the mobile device with highly limited computational resources. Classification of iris sample for spoof and live is made by a single frame using pair of images: eye region and normalized iris. The following types of iris spoof samples are considered in this particular work: printed on paper, printed on paper with imposition of a contact lens, printed on paper with application of transparent glue. Testing of the method is performed on different datasets, both open and manually collected, and revealed its high performance on both the classification accuracy and processing speed. The method also showed its robustness under uncontrollably changing environmental conditions, which are specific and significant when interacting with the mobile device.

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

- [1] *Efimov I.* Iris Antispoofing Solution for Mobile Biometric Applications // Pattern Recognition and Image Analysis, Moscow: Pleiades Publishing, Ltd., 2018 (in press).

Выделение отличий на разносезонных изображениях земной поверхности при помощи генеративных состязательных сетей

*Лебедев Максим Алексеевич*¹★

lebedev_maxim@list.ru

*Комаров Денис Валерьевич*²

mrkomap@mail.ru

*Выголов Олег Вячеславович*³

o.vygodov@gosniias.ru

*Визильтер Юрий Валентинович*⁴

viz@gosniias.ru

^{1,2,3,4}Москва, Государственный научно-исследовательский институт авиационных систем

DOI: 10.30826/IDP201859

Анализ изменений земной поверхности в разновременных последовательностях аэрофотоснимков является важной частью многих приложений, таких, например, как анализ развития городской застройки, анализ последствий стихийных бедствий, сельскохозяйственный мониторинг и т.д. Ручная обработка снимков земной поверхности трудоемкая задача, занимающая много времени, поэтому автоматическое обнаружение изменений на аэрофотоснимках является важной областью исследований.

В данной работе предлагается метод выделения отличий на разносезонных аэрофотоснимках при помощи генеративных состязательных сетей. Предлагается архитектура сети для построения карты отличий на основе оригинальной архитектуры pix2pix. В работе описываются результаты трех последовательных исследований: выделение отличий на модельных изображениях без относительного сдвига объектов, выделение отличий на модельных изображениях в случае небольшого относительного сдвига объектов и выделение отличий на реальных разносезонных изображениях земной поверхности. Приводятся результаты работы предложенного метода на общедоступном наборе данных по выделению отличий AICDDataset.

Работа выполнена при поддержке РФФИ, грант 16-11-00082.

- [1] *Lebedev, M. A., Vizilter, Y. V., Vygodov, O. V., Knyaz, V. A., and Rubis, A. Y.* Change detection in remote sensing images using conditional adversarial networks, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2, 565-571 c., 2018. <https://doi.org/10.5194/isprs-archives-XLII-2-565-2018>

Change detection in remote sensing images using conditional adversarial networks

*Lebedev Maksim*¹★

lebedev_maxim@list.ru

*Komarov Denis*²

mrkomap@mail.ru

*Vygodov Oleg*³

o.vygodov@gosniias.ru

*Vizilter Yuri*⁴

viz@gosniias.ru

^{1,2,3,4}Moscow, State Research Institute of Aviation Systems

Change detection in the time-varying sequences of remote sensing images acquired on the same geographical area is an important part of many practical applications, e.g. urban development analysis, environmental inspection, agricultural monitoring. In most cases, solving the change detection task in manual mode is a highly time-consuming operation, which makes an automation of this process an important and practically demanded field of research.

We present a method for change detection in season-varying remote sensing images using Generative Adversarial Network approach. The original network architecture based on pix2pix is proposed and evaluated for difference map creation. The paper address three types of experiments: change detection in synthetic images without objects relative shift, change detection in synthetic images with small relative shift of objects, and change detection in real season-varying remote sensing images. The results of the proposed method on the publicly available data set for change detection AICDDataset are presented.

This work was performed with the support of Russian Science Foundation (RNF grant 16-11-00082).

- [1] *Lebedev, M. A., Vizilter, Y. V., Vygodov, O. V., Knyaz, V. A., and Rubis, A. Y.* Change detection in remote sensing images using conditional adversarial networks, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2, 565-571 p., 2018. <https://doi.org/10.5194/isprs-archives-XLII-2-565-2018>

Сравнение информативных признаков изображений радужки глаза методом оптимального пути

*Новик Владимир Петрович*¹

novikvp@mail.ru

Матвеев Иван Алексеевич^{1,2*}

matveev@ccas.ru

¹Москва, Iritech Inc.

²Москва, ФИЦ ИУ РАН

DOI: 10.30826/IPD201860

В большинстве методов распознавания радужной оболочки глаза используются локальные текстурные признаки. Качество распознавания существенно зависит от точности их совмещения при сравнении. Предложен простой в сравнении с существующими аналогами, при этом столь же эффективный метод повышения точности совмещения, основанный на поиске оптимального пути в пространстве смещений локальных участков радужки. Исследуется вопрос выбора оптимальных параметров вейвлетов, формирующих шаблон радужной оболочки глаза. Показано, что улучшение точности совмещения позволяет использовать более информативные высокочастотные вейвлеты. Численные эксперименты проведены на базах данных изображений ICE2005 и CASIA, находящихся в свободном доступе.

Работа поддержана грантом РФФИ № 16-07-01171.

- [1] *Новик В. П., Матвеев И. А., Леонов В. Ю.* Сравнение информативных признаков изображений радужки глаза методом оптимального пути // Информационные технологии, Москва: Новые технологии, 2018. — N. 10. (в печати) .

Matching of informative features of iris images with the optimal path method

*Novik Vladimir*¹

novikvp@mail.ru

Matveev Ivan^{1,2}★

matveev@ccas.ru

¹Moscow, Iritech Inc.

²Moscow, FRC CSC RAS

The majority of iris recognition methods use local textural features. The accuracy then depends significantly on the precision of local textural features alignment. A simple (in comparison with state of the art) yet efficient method of enhancing the alignment accuracy is proposed. It is based on locating the optimal path in the space of displacements of local iris patterns. The problem of choosing the optimal parameters of wavelets forming the iris template is studied. It is shown that improvement of alignment accuracy allows using more informative high-frequency wavelets. Computational experiments were carried out with ICE2005 and CASIA bases from public domain.

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

- [1] *Novik V., Matveev I., Leonov V.* Matching of informative features of iris images with the optimal path method // Information technologies, Moscow: Novye tehnologii, 2018. N.10 (in press).

Теоретические основы гипертрейс–преобразования

*Федотов Николай Гаврилович*¹ fedotov@pnzgu.ru

*Сёмов Алексей Александрович*² matematik_aleksey@mail.ru

Моисеев Александр Владимирович^{1*} moigus@mail.ru

¹Пенза, ФГБОУ ВО «Пензенский государственный университет»

²Пенза, ООО «Комэрф»

DOI: 10.30826/IDP201861

В статье описывается теоретическая основа нового геометрического метода анализа и распознавания трехмерных (3D) изображений. Приведена техника сканирования для формирования гипертрейс–преобразования и его математическая модель. Данный метод в отличие от существующих позволяет анализировать 3D–изображения без предварительного их упрощения, анализируя непосредственно их трехмерную форму. Обоснован выбор сканирующего инструмента и построение опорной сетки на сфере, необходимой для решения проблемы инвариантности распознавания 3D–изображения к повороту. Разработан математический аппарат реализации техники сканирования на основе стохастической геометрии и функционального анализа. Введен новый математический инструмент для анализа 3D–изображений — гипертрейс–матрица, позволяющий распознавать пространственные объекты сложной формы и структуры благодаря построению единой математической модели. Дано описание нового типа признаков 3D–изображений, имеющих аналитическую структуру, — гипертриплетные признаки, благодаря аналитической структуре которых возможна автоматическая генерация большого количества признаков с заранее заданными свойствами. Рассматриваются результаты проверки, демонстрирующие точность вычисления признаков для распознавания 3D–изображения и доказывающие адекватность разработанного математического аппарата.

- [1] *Федотов Н. Г.* Теоретические основы гипертрейс–преобразования: техника сканирования, математический аппарат и экспериментальная проверка / Н. Г. Федотов, А. А. Сёмов, А. В. Моисеев // Компьютерная оптика, Самара: ИСОИ РАН, 2018. Т.42, №2. — С. 273–282. — DOI: 10.18287/2412-6179-2018-42-2-273-282. <http://www.computeroptics.smr.ru/KO/PDF/KO42-2/420213.pdf>.

Theoretical Foundations of Hypertrace–transform

*Fedotov Nikolai*¹

fedotov@pnzgu.ru

*Syemov Alexey*²★

mathematik.aleksey@mail.ru

*Moiseev Alexander*¹★

moigus@mail.ru

¹Penza, Penza State University

²Penza, LLC Comearth

We consistently describe the theoretical basis of a new geometric method of analysis and recognition of three-dimensional (3D) images. The description of a scanning technique for forming a hypertrace transform and its mathematical model are given. This method, unlike the existing ones, enables 3D images to be analyzed directly from their 3D shape, without first simplifying them or constructing plane projections. We substantiate the selection of a particular scanning tool and the need to construct a reference spherical grid to address the problem of the rotational invariance of the 3D image recognition. A mathematical apparatus of the stochastic realization of the scanning technique based on stochastic geometry and functional analysis is developed. We introduce a new mathematical tool for 3D image analysis — a hypertrex matrix that allows spatial objects of complex shape and structure to be recognized by constructing a single mathematical model of the 3D image. We describe a new type of 3D image features that have an analytic structure — hypertryplet features, whose analytical structure makes possible an automatic generation of a large number of features with predetermined properties. Results of the experimental verification are presented, demonstrating the accurate calculation of features for 3D image recognition and proving the adequacy of the developed mathematical apparatus.

- [1] *Fedotov NG, Syemov AA, Moiseev AV*. Theoretical foundations of hypertracettransform: scanning techniques, mathematical apparatus and experimental verification // Computer Optics, Samara: IPSI RAS — Branch of the FSRC “Crystallography and Photonics” RAS, 2018; 42(2): 273-282. — DOI: 10.18287/2412-6179-2018-42-2-273-282. <http://www.computeroptics.smr.ru/KO/PDF/K042-2/420213.pdf>.

Анализ эффективности нелинейной фильтрации многозональных спутниковых изображений и их последовательностей

*Васильев Константин Константинович*¹ vkk@ulstu.ru

*Дементьев Виталий Евгеньевич*¹ dve@ulntc.com

Андрьянов Никита Андреевич^{1,2*} nikita-and-nov@mail

¹Ульяновск, Россия, Ульяновский государственный технический университет

²Ульяновск, Россия, Ульяновский институт гражданской авиации имени Главного маршала авиации Б. П. Бугаева

DOI: 10.30826/IDP201862

В работе рассмотрена задача представления сложных и неоднородных многозональных изображений с помощью дважды стохастических авторегрессионных моделей случайных полей. Для повышения качества описания реальных снимков предложено использовать дважды стохастическую модель на базе авторегрессий с кратными корнями характеристических уравнений. Синтезируется алгоритм оценивания изображений и их последовательностей с использованием нелинейного фильтра Калмана. Разрабатываемый фильтр представлен в матричной форме, что позволяет вести обработку многомерных изображений. Проводится сравнение с известными алгоритмами локальной фильтрации и алгоритмами на базе авторегрессий. Получены характеристики эффективности фильтрации последовательности спутниковых изображений в зависимости от длины последовательности и коэффициента межкадровой корреляции. Получены выигрыши по дисперсии ошибки фильтрации, составляющие в среднем 25-30% [1].

Работа поддержана грантами РФФИ №№ 17-01-00179 и 16-41-732027.

- [1] *K. Vasiliev , V. Dementiev and N. Andriyanov* Representation and processing of multispectral satellite images and sequences // *Procedia Computer Science* 126 (2018), p. 49-58 https://ac.els-cdn.com/S1877050918311827/1-s2.0-S1877050918311827-main.pdf?_tid=1733dba1-18dc-4183-9f2c-6cc78fcbf66e&acdnat=1536635582_f339a7b54c16c0fe960da7ddbfe32452.

Analysis of nonlinear filtering efficiency in processing multizone satellite images and their sequences

*Vasiliev Konstantin*¹

vkk@ulstu.ru

*Dementiev Vitaliy*¹

dve@ulntc.com

Andriyanov Nikita^{1,2*}

nikita-and-nov@mail

¹Russia, Ulyanovsk, Ulyanovsk State Technical University

²Russia, Ulyanovsk, Ulyanovsk Institute of Civil Aviation

The problem of representation of complex and inhomogeneous multizone images using the doubly stochastic autoregressive models of random fields is considered. To improve the quality of the description of real images, it is proposed to use a doubly stochastic model based on autoregressions with multiple roots of the characteristic equations. An algorithm for estimating images and their sequences using a nonlinear Kalman filter is synthesized. The filter being developed is presented in matrix form, which allows processing multidimensional images. Comparison with known local filtering algorithms and algorithms based on autoregressions is carried out. The characteristics of the efficiency of filtering the sequence of satellite images are obtained depending on the length of the sequence and the coefficient of interframe correlation. The gains are obtained for dispersion of the filtering error, which averaged 25-30% [1].

This research is funded by the Russian Foundation for Basic Research, grants 17-01-00179 and 16-41-732027.

- [1] *K. Vasiliev , V. Dementiev and N. Andriyanov* Representation and processing of multispectral satellite images and sequences // *Procedia Computer Science* 126 (2018), p. 49-58 https://ac.els-cdn.com/S1877050918311827/1-s2.0-S1877050918311827-main.pdf?_tid=1733dba1-18dc-4183-9f2c-6cc78fcfb66e&acdnat=1536635582_f339a7b54c16c0fe960da7ddbfe32452.

Метод выделения радужки на изображении, полученном в сложных условиях окружения

Одиноких Глеб Андреевич^{1*} g.odinokikh@gmail.com

*Коробкин Михаил Владимирович*² m.korobkin@hotmail.com

*Ефимов Юрий Сергеевич*³ yuri.efimov@phystech.edu

*Соломатин Иван Андреевич*³ ivan.solomatin@phystech.edu

Матвеев Иван Алексеевич^{1,3} matveev@ccas.ru

¹ г.Москва, ФИЦ ИУ РАН

² г.Зеленоград, НИУ МИЭТ

³ г.Долгопрудный, МФТИ

DOI: 10.30826/IDP201863

Выделение (сегментация) области радужки на изображении - один из основных этапов распознавания. Ошибки сегментации влекут за собой рост ошибок, делая систему менее надежной и удобной в использовании. Подавляющее большинство существующих подходов ориентированы на использование систем в условиях слабо изменяющегося окружения. Классические методы, основанные на эвристиках, хорошо зарекомендовали себя здесь. Широкое распространение технологий распознавания создает необходимость обеспечения полной функциональности систем в более экстремальных условиях и, как следствие, создания более гибких и устойчивых решений. В данной работе рассматривается применение CNN для выделения радужки в условиях, максимально приближенных к реальным сценариям использования. В работе предложены новые архитектуры сетей, сочетающие в себе последние достижения в области семантической сегментации и классификации изображений. Архитектуры FCN и SegNet были взяты за основу и "усилены" остаточными блоками. Предложенный подход позволил понизить требования к производительности целевой платформы, а так же достигнуть 0.93 и 0.92 IoU на оригинальном и модифицированном наборах данных CASIA-Iris-Lamp, что является значительным улучшением в сравнении с существующими методами. Работа поддержана грантом РФФИ № 16-07-01171.

- [1] *Korobkin M., Odinokikh G.* Iris Segmentation in Challenging Conditions // Proc. of IAPRAI, Montreal: CENPARMI, 2018, 656–660. <https://users.encs.concordia.ca/~icprai18/ICPRAI%202018%20Proceedings.pdf>.

Iris Segmentation in Challenging Conditions

*Odinokikh Gleb*¹*

g.odinokikh@gmail.com

*Korobkin Mikhail*²

mikhail.korobkin@hotmail.com

*Efimov Yuri*³

yuri.efimov@phystech.edu

*Solomatin Ivan*³

ivan.solomatin@phystech.edu

Matveev Ivan^{1,3}

matveev@ccas.ru

¹Moscow, FRC CSC RAS

²Zelenograd, MIEE

³Dolgoprudny, MIPT

Iris segmentation is an irreplaceable stage of iris recognition pipeline. Wrong segmentation decreases recognition performance and makes the technology inconvenient for users. Previously when conditions were mild and controlled the task was solved by image processing techniques and rule based approaches. Nowadays widespread of biometric technologies has relaxed operation conditions for such systems demanding more flexible and robust solutions.

Deep learning and convolutional neural networks represent a class of machine learning methods that secured top results in many computer vision tasks. The latest contribution to iris segmentation was also made by neural networks. In spite of previously achieved great results this work addresses even more challenging conditions that allows iris recognition to be used in wide range of real life cases. Novel CNN architectures are proposed in this work. They were designed to combine the latest achievements in classification and semantic segmentation fields. FCN and SegNet architectures have been picked up as prototypes and were strengthened by residual blocks. This allowed to make lightweight networks that could be shipped on various embedded platforms to successfully operate under less controllable environmental conditions. The approach allowed to obtain 0.93 and 0.92 IoU on original and modified CASIA-Iris-Lamp datasets which is a significant improvement in comparison with the results achieved before.

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

- [1] *Korobkin M., Odinokikh G.* Iris Segmentation in Challenging Conditions // Proc. of IAPRAI, Montreal: CENPARMI, 2018, 656–660. <https://users.encs.concordia.ca/~icprai18/ICPRAI%202018%20Proceedings.pdf>.

Использование Microsoft Kinect V2 для детектирования падений человека

Середин Олег Сергеевич

oseredin@yandex.ru

Копылов Андрей Валериевич

and.kopylov@gmail.com

*Родионов Денис Сергеевич**

den191097@gmail.com

Тула, Тульский государственный университет

DOI: 10.30826/IDP201864

Модуль своевременного и надежного обнаружения падений человека является одной из ключевых компонент систем удаленного мониторинга за состоянием пожилых людей. Вместе с тем, камеры глубины занимают все большее место в системах распознавания событий, поскольку они стали доступнее. Некоторые из них позволяют непосредственно получить скелетное описание фигуры человека. Возможность получения точек скелетного описания по карте глубины дает способ компактного представления позы человека и предотвращает выход детальной информации о глубине или исходного видео за пределы системы, что позволяет поднять приватность всей системы. В исследовании на основе сравнения различных RGB-D камер, выбрана наиболее перспективная для дальнейших разработок модель - Microsoft Kinect v2. В работе в качестве базы для экспериментов используется TST Fall Detection Dataset v2. Предложен алгоритм на основе признаков кодирования скелета на последовательности соседних кадров, используется классификатор на основе метода опорных векторов и процедуры на основе кумулятивной суммы для согласования решений в последовательности кадров. Для отбраковки некачественных скелетов предложено использовать одноклассовый классификатор. Получено качество детектирования падения на уровне 0.95 для процедуры кросс-проверки на основе удаления записей конкретной персоны из базы. Работа поддержана грантом РФФИ № 16-57-52042, 18-07-00942.

- [1] *Seredin O., Kopylov A., Rodionov D.* A skeleton features-based fall detection using Microsoft Kinect v2 with one class-classifier outlier removal // IEEE Transactions on Pattern Analysis and Machine Intelligence. Special Issue on RGB-D Vision: Methods and Applications, на рецензировании. <http://lda.tsu.tula.ru/publication/>.

Microsoft Kinect v2 for fall detection

Seredin Oleg

oseredin@yandex.ru

Kopylov Andrei

and.kopylov@gmail.com

*Rodionov Denis**

den191097@gmail.com

Tula, Tula State University

The module of real-time and robust fall detection is one of the key components of elderly people care and monitoring systems. At the same time, depth sensors, as they became more available, occupy an increasing place in event recognition systems. Some of them can directly produce a skeletal description of the human figure. The ability to obtain skeletal descriptions from the depth map provides a way of compact representation of a person's posture and prevents the output of source video or detailed information about the depth outside the system, which allows raising the privacy of the entire system. Based on a comparative study of different RGB-d cameras, the most promising model for further development was chosen - Microsoft Kinect v2. The TST Fall Detection Dataset v2 with records of 11 persons is used in the work as a base for experiments. The proposed algorithm is based on the skeleton features encoding on the sequence of neighboring frames and support vector machine classifier. A version of a cumulative sum method is applied for combining the individual decisions on the consecutive frames. It is offered to use the one-class classifier for detection of low-quality skeletons. The 0.95 accuracy of our fall detection procedure was obtained in the cross-validation procedure based on the removal of records of a particular person from the database (Leave-one-Person-out).

This research is funded by RFBR, grants 16-57-52042, 18-07-00942.

- [1] *Seredin O., Kopylov A., Rodionov D.* A skeleton features-based fall detection using Microsoft Kinect v2 with one class-classifier outlier removal // IEEE Transactions on Pattern Analysis and Machine Intelligence. Special Issue on RGB-D Vision: Methods and Applications, under review. <http://lda.tsu.tula.ru/publication/>.

Аппроксимация границ радужки классифицирующей свёрточной нейронной сетью

*Соломатин Иван Андреевич*¹★ ivan.solomatin@phystech.edu

*Одиноких Глеб Андреевич*² g.odinokikh@gmail.com

*Ефимов Юрий Сергеевич*¹ yuri.efimov@phystech.edu

*Матвеев Иван Алексеевич*² matveev@ccas.ru

¹Москва, МФТИ (ГУ)

²Москва, ФИЦ ИУ РАН

DOI: 10.30826/IDP201865

Идентификация человека по радужной оболочке глаза — актуальная задача в биометрических системах. Для выполнения идентификации необходимо найти область локализации радужки на изображении. Первый этап её локализации — выделение области глаза на изображении. Затем, в этой области находятся границы радужки, которые аппроксимируются некоторыми кривыми, обычно — окружностями.

В данной работе описан алгоритм, основанный на классифицирующей свёрточной нейронной сети, позволяющий получить параметры аппроксимации границ радужки окружностями с ошибкой не более 5% от диаметра радужки. Задача нахождения каждого из параметров аппроксимации (координаты центров и радиусы окружностей, задающих внутреннюю и внешнюю границу радужки) сводится к задаче классификации. Считается, что на квадратном входном изображении $w \times w$ полностью содержится глаз, причём диаметр внешней границы радужки не менее $\frac{1}{4}w$.

Точность работы алгоритма была измерена на нескольких открытых базах изображений радужной оболочки глаза (CASIA3-Iris-Lamp, CASIA-Iris-M1, ICE DB), а так же на базе, собранной автором с помощью микрокомпьютера Raspberry pi. Средняя точность алгоритма на вышеописанных базах составила $\sim 99\%$.

Для обучения нейронной сети была использована библиотека keras для языка python. Алгоритм так же был реализован на языке C.

Работа поддержана грантом РФФИ №16-07-01171.

- [1] *M.Korobkin, G.Odinokikh, Y.Efimov, I.Solomatin and I.Matveev* Iris Segmentation in Challenging Conditions // Pattern Recognition and Image Analysis, Springer, 2018 (в печати).

Iris boundaries approximation by classifying convolutional neural network

Solomatin Ivan^{1*}

ivan.solomatin@phystech.edu

*Odinokikh Gleb*²

g.odinokikh@gmail.com

*Efimov Yuri*¹

yuri.efimov@phystech.edu

*Matveev Ivan*²

matveev@ccas.ru

¹Moscow, MIPT

²Moscow, FRC CSC RAS

Person identification by the iris image is an actual problem in biometric systems. To perform the identification, it is necessary to find the region of iris localization in the image. The first stage of this localization is finding eye region on the image. In this region there are iris inner and outer boundaries, which are usually approximated by circles.

This paper describes an algorithm based on a classifying convolutional neural network, which obtains parameters for approximating iris boundaries by circles with an error of no more than 5% of the iris diameter. The problem of finding each of the approximation parameters (coordinates of centers and radii of circles defining the inner and outer boundaries of the iris) is solved as classification problem. Input image is supposed to be a square $w \times w$ and to contain entire iris. Outer iris boundary diameter is supposed to be not less than $\frac{1}{4}w$.

The accuracy of the algorithm was evaluated on several open iris image databases (CASIA3-Iris-Lamp, CASIA-Iris-M1, ICE DB), as well as on a database which was collected by the author using a Raspberry pi microcomputer. The average accuracy of the algorithm on these databases was $\sim 99\%$.

To train the neural network, the keras library for the python language was used. The algorithm was also implemented on C.

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

- [1] *M.Korobkin, G.Odinokikh, Y.Efimov, I.Solomatin and I.Matveev* Iris Segmentation in Challenging Conditions // Pattern Recognition and Image Analysis, Springer, 2018 (in press).

Мера TF-IDF, сила связи слов и эталон языкового описания единицы знаний в открытых тестах

Емельянов Геннадий Мартинович Gennady.Emelyanov@novsu.ru

*Михайлов Дмитрий Владимирович** mdv74@list.ru

Козлов Александр Павлович caleo@yandex.ru

Великий Новгород, Россия, НовГУ

DOI: 10.30826/IDP201866

Разработка открытых тестов предполагает накопление и систематизацию экспертных знаний, исходно представляемых текстами естественного языка. Конечная цель — найти наиболее рациональный вариант передачи смысла для множества семантически эквивалентных форм его выражения. Данная задача в работе решается во взаимосвязи с поиском составляющих образа исходной фразы (СОИФ) в текстах тематического корпуса. При этом основой выделения СОИФ является совместное использование оценки силы связи сочетаний её слов, встречающихся во фразах анализируемого текста, и разбиения этих слов на классы по значению TF-IDF относительно текстов корпуса. Расширением связей слов до трёх и более элементов (с привлечением базы известных синтаксических отношений и без неё) вводится оценка релевантности корпуса единице знаний по степени охвата слов исходной фразы наиболее значимыми совокупностями связей относительно документов, в которых СОИФ представлены наиболее полно. Предложено использование данной оценки для целенаправленного отбора из текстов корпуса фраз, взаимно эквивалентных либо дополняющих друг друга по смыслу. Для ранжирования фраз по близости смысловому эталону вводятся три альтернативные оценки: на основе классификации слов исходной фразы по значению TF-IDF и на основе оценки силы их связи (с учётом предлогов и союзов и без них). При этом достигается минимум двукратное сокращение текстовой информации, необходимой для представления выделяемой единицы знаний. Работа поддержана грантом РФФИ № 16-01-00004.

- [1] *Емельянов Г. М., Михайлов Д. В., Козлов А. П.* Релевантность множества тематических текстов единице знаний и оценка близости языковых форм её выражения смысловому эталону // *Распознавание образов и анализ изображений*, 2018 (в печати).

TF-IDF, coupling strength of words and standard for lingual description of knowledge unit in open tests

Emelyanov Gennady

Gennady.Emelyanov@novsu.ru

*Mikhaylov Dmitry**

mdv74@list.ru

Kozlov Aleksandr

caleo@yandex.ru

Russia, Veliky 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 texts. The main goal is finding the most rational variant to express the sense for the set of semantically equivalent forms of its expression in a given human language. In current paper this problem is solved in interrelation with the search of initial phrase's image constituents (IPhICs) in texts of topical corpus. Here a basis for selection of IPhICs be the estimation of coupling strength of words from initial phrase jointly occurring in phrases of the analyzed text together with classifying these words according to TF-IDF metrics in relation to corpus texts. Herewith the extension of links of words from bigrams to three and more elements allows to estimate the relevance of text corpus to knowledge unit defined by initial phrase or their set using the coverage degree of words of initial phrase by the most significant sets of links concerning the documents which are the best in reflection of IPhICs. Variants of link revelation with and without application of a database of known syntactical relations are considered with regard to this relevance estimation. It was discussed how to use the offered estimation for the purposeful selection in corpus texts the phrases mutually equivalent or complementary in sense. For ranking these phrases according to the affinity to sense standard three alternative estimations were proposed: basing on clustering of words of initial phrase according to their TF-IDF and on the base of coupling strength of these words. Herewith at least a two-fold reduction in the textual information necessary to represent the extracted knowledge unit is achieved. This research is funded by RFBR, grant 16-01-00004.

- [1] *Emelyanov G., Mikhaylov D., Kozlov A.* 2018 (in press). Relevance of textual set to knowledge unit and estimation of affinity to sense standard for its linguistic expressional means. *Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications.*

Тематический информационный поиск

Янина Анастасия Олеговна *

yanina@phystech.edu

Воронцов Константин Вячеславович

voron@forecsys.ru

Москва, Лаборатория машинного интеллекта МФТИ

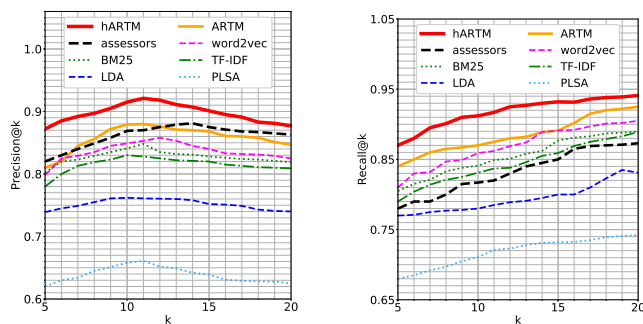
DOI: 10.30826/IDP201867

Тематическое моделирование позволяет представлять тексты естественного языка неотрицательными разреженными векторами низкой размерности и искать документы, близкие по смыслу к тексту запроса. В отличие от обычного поиска, запросы могут быть произвольной длины.

В докладе рассматривается серия моделей, в которых комбинирование иерархического разделения тем на подтемы (hARTM), разреживающих регуляризаторов и модальностей авторов, категорий, тегов и биграмм (ARTM) превосходит по критериям точности и полноты первых k позиций поисковой выдачи не только полнотекстовые методы (TF-IDF, BM25) и латентные модели (PLSA, LDA, word2vec), но и результаты работы ассессоров.

Оценивание качества поиска проводилось на коллекциях технологических новостей habr.com и techcrunch.com с помощью краудсорсинга по 100 поисковым заданиям. Каждое задание независимо выполняли три ассессора. При этом ассессоры тратят на каждое задание в среднем около 30 минут, тогда как тематический поиск выдаёт более качественный результат менее чем за секунду.

Работа поддержана грантом РФФИ № 17-07-01536.



- [1] Yanina A., Golytsin L., Vorontsov K. Multi-objective topic modeling for exploratory search in tech news // CCIS, vol.789, Springer, 2018.

Topic-based information retrieval

Anastasia Ianina ^{*}
Konstantin Vorontsov

yanina@phystech.edu
voron@forecsys.ru

Moscow, MIPT Machine Intelligence Laboratory

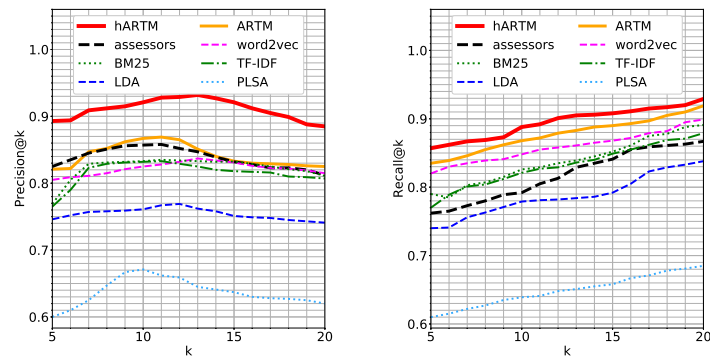
Topic modeling produces a non-negative low-dimensional sparse vector representation of texts. Topic-based information retrieval uses these representations to find semantically similar documents for long text queries.

We present a bunch of models based on hierarchical topic modeling (hARTM), sparsing regularizers, and additional modalities including authors, categories, tags, and bigrams. Combining all these features together, we obtain a topic-based search engine that outperforms not only common full-text search methods (TF-IDF, BM25) or latent models (PLSA, LDA, word2vec), but also human assessors.

In our experiments we use two tech news collections: habr.com (in Russian) and techcrunch.com (in English). We evaluate the models with precision and recall at top- k search result positions. Human rankings are obtained through crowdsourcing for 100 search tasks, where each task is independently processed by 3 assessors.

We observe that assessors spend about 30 minutes on each task on average, while the topic-based search yields better quality results in a second.

This work was supported by RFBR grant 17-07-01536.



- [1] Ianina A., Golytsin L., Vorontsov K. Multi-objective topic modeling for exploratory search in tech news // CCIS, vol.789, Springer, 2018.

Гиперграфовые многомодальные вероятностные тематические модели транзакционных данных

Жариков Илья Николаевич * zharikov.i.n@yandex.ru

Апишев Мурат Азаматович great-mel@yandex.ru

Воронцов Константин Вячеславович voron@forecsys.ru

Москва, Московский Физико-Технический Институт

DOI: 10.30826/IDP201868

Вероятностная тематическая модель коллекции текстовых документов строит распределения слов и документов по кластерам-темам. Тематическая модель данных рекомендательной системы распределяет множества пользователей и объектов по кластерам-интересам. В этих задачах выборка данных описывает парные взаимодействия «слова–документы» и «субъект–объект» соответственно. В общем случае можно говорить, что выборка состоит из рёбер двудольного графа. В данной работе предлагается аддитивно регуляризованная тематическая модель (ARTM) для более общего случая, когда каждое взаимодействие (транзакция) происходит между тремя или более объектами различных модальностей. В этих случаях выборка представляется множеством рёбер гиперграфа. Регуляризованный EM-алгоритм TransARTM для таких случаев реализован нами в проекте с открытым кодом BigARTM.org. Такие задачи возникают при анализе данных социальных медиа, рекламных и торговых сетей, рекомендательных систем и банков. В текстах объединение слов предложения в одну транзакцию соответствует лингвистической гипотезе, что каждое высказывание естественного языка относится только к одной теме. В экспериментах на модельных данных TransARTM оказался более устойчивым, быстрее сходился и демонстрировал лучшие результаты на малых выборках по сравнению с обычными мультимодальными тематическими моделями. На реальных данных музыкальной рекомендательной системы TransARTM показал значительно лучшие результаты, чем базовая модель, основанная на рекомендации наиболее популярных треков.

- [1] *Жариков И. Н.* Многомодальные тематические модели на гиперграфах. МФТИ, 2018.
- [2] *Воронцов К. В.* Вероятностное тематическое моделирование: обзор моделей и аддитивная регуляризация. 2018.

Hypergraph multimodal probabilistic topic modeling of transaction data

Ilya Zharikov *

zharikov.i.n@yandex.ru

Murat Apishev

great-mel@yandex.ru

Konstantin Vorontsov

voron@forecsys.ru

Moscow, Moscow Institute of Physics and Technology

Probabilistic topic models of text documents infer distributions of words and documents over topics. Similarly, probabilistic topic models for recommender systems infer distributions of users and items over topics. The datasets of these tasks consist of “word–document” and “user–item” transactions respectively. More formally, in these and other similar cases the input data can be represented by a set of edges of a bipartite graph.

In this work we consider a more general case, where a transaction may be formed by three or more objects of arbitrary modalities. This corresponds to a set of edges of a hypergraph as the input. We generalize the framework of additively regularized topic models (ARTM) for this case and implement a transactional EM-like algorithm, that we call TransARTM. It is now available in the open source project BigARTM.org.

Hypergraph topic models can be applied in many domains, including social media, advertising networks, recommender systems, and banking. In texts, considering all words of a phrase as one transaction corresponds to a linguistic assumption that the whole phrase refers only to one topic.

In our experiments on synthetic transactional data, TransARTM demonstrated greater stability, faster convergence, and better results for small datasets if compared to conventional multimodal topic models. On the real data of the music recommender system, TransARTM showed significantly better results if compared to a baseline that recommends the most popular tracks.

- [1] *Zharikov I.* Multimodal topic modeling on hypergraphs. MIPT, 2018.
- [2] *Vorontsov K.* Probabilistic topic modeling: the survey of models and additive regularization. 2018.

Формализация задач поиска сгущений в разреженных булевых матрицах

Алешин Илья Сергеевич

ilyaaln@yandex.ru

Россия, Москва, МГУ ВМК

DOI: 10.30826/IDP201869

В огромном количестве прикладных задач интеллектуального анализа данных, таких как исследование геной экспрессии и тканей, текстовой и веб-информации, рыночных корзин, клиентских сред, входная информация естественным образом представляется в виде двумерной матрицы «субъекты-объекты» («клиенты-сервисы»). Основной целью в рамках указанных областей является так называемая бикластеризация данных, то есть выделение групп в определённом смысле схожих строк и столбцов. Немалая часть таких задач характеризуется сильной разреженностью соответствующих матриц. Важным аспектом бикластеризации является поиск в некотором смысле плотных подматриц в булевых матрицах, что и является основной целью данного исследования. Понятия разреженности и плотности являются плохо формализованными, что делает целесообразным применение методов алгебраического подхода для поиска таких подматриц. В работе предложено описание предметной области в рамках алгебраического подхода, приведены системы универсальных и локальных ограничений, сформулированы и доказаны соответствующие критерии разрешимости рассматриваемых задач.

Работа поддержана грантом РФФИ № 17-20-02200.

- [1] *Алешин И. С.* О формальной постановке задач поиска сгущений в разреженных булевых матрицах // Информатика и её применение, 2018, Том 12 №1, стр. 40–48 // http://www.ipiran.ru/journal/issues/2018_12_01/Vol12_Issue1.pdf.

Formalization of tasks searching dense submatrices in boolean sparse matrices

Aleshin Ilya

ilyaaln@yandex.ru

¹Moscow, Russia, MSU CMC

In a significant part of data mining applications such as microbiology, gene expression data, text and web information, market baskets, customer environments, input information is represented as a two-dimensional matrix $(\text{subjects-objects})_{i,j}$ ($\text{clients-services})_{i,j}$). A great quantity of such problems is characterized by strong sparse data. Therefore, the important part of such problems is the so-called biclustering of data, based on the selection of homogeneous groups of rows and columns. In this paper we consider the boolean sparse matrices. The main goal of our research is to search in some sense dense submatrices within the theory framework of the algebraic approach. The concepts of sparsity and density are poorly formalized, which makes expedient to use methods of the algebraic approach for the search such submatrices. In this paper subject area is described, the systems of universal and local constraints are proposed, the corresponding criteria for the solvability of the problems under consideration are formulated and proved.

This research is funded by RFBR, grant 17-20-02200.

- [1] *Aleshin Ilya* On the formalization of tasks searching dense submatrices in boolean sparse matrices // Informatics and applications, 2018 volume 12 issue 1.

Модель классификации на основе средней взаимной информации

Ланге М.М.^{1*}

lange_mm@ccas.ru

Ланге А.М.¹

lange_am@mail.ru

¹Москва, ФИЦ ИУ РАН

DOI: 10.30826/IDP201870

Предлагается теоретико-информационная модель классификации данных на основе схемы кодирования сообщений с заданной точностью, переданных по каналу с шумом. Пусть $\Omega^c = \{\omega_1, \dots, \omega_c\}$, $c \geq 2$ – множество классов (алфавит сообщений) с априорными вероятностями $P(\omega_i) > 0$, $i = 1, \dots, c$ и \mathbf{X} – множество объектов на выходе канала наблюдения с условными по классам плотностями распределения $\{p(\mathbf{x}|\omega_i), i = 1, \dots, c\}$, $\mathbf{x} \in \mathbf{X}$. Множество решений $\Omega = \Omega^c$ по объектам из \mathbf{X} формируется на выходе тест-канала, заданного условными вероятностями решений $\mathbf{Q} = \{Q(\omega_j|\mathbf{x}), j = 1, \dots, c\}$, $\omega_j \in \Omega$, по объекту $\mathbf{x} \in \mathbf{X}$. Тройка $\Omega^c, \mathbf{X}, \Omega$, порождает модель классификации

$$\Omega^c \Rightarrow \boxed{\{p(\mathbf{x}|\omega_i), i = 1, \dots, c\}} \Rightarrow \mathbf{X} \Rightarrow \boxed{\{Q(\omega_j|\mathbf{x}), j = 1, \dots, c\}} \Rightarrow \Omega$$

в которой средняя взаимная информация $I_{\mathbf{Q}}(\mathbf{X}; \Omega) \geq 0$ и вероятность ошибки $E_{\mathbf{Q}}(\mathbf{X}, \Omega) > 0$ зависят от тест-канала \mathbf{Q} . Для принятой модели вводится условный минимум

$$R(\varepsilon) = \min_{\mathbf{Q}: E_{\mathbf{Q}}(\mathbf{X}, \Omega) \leq \varepsilon} I_{\mathbf{Q}}(\mathbf{X}; \Omega)$$

по \mathbf{Q} , который дает наименьшую вероятность ошибки $\varepsilon > 0$ при фиксированной средней взаимной информации, определяющей вычислительную сложность поиска решения.

Получена нижняя граница $R_L(\varepsilon) \leq R(\varepsilon)$ в форме монотонно убывающей функции

$$R_L(\varepsilon) = I(\mathbf{X}; \Omega^c) - h(\varepsilon - \varepsilon_{\min}) - (\varepsilon - \varepsilon_{\min}) \ln(c-1), \quad \varepsilon_{\min} \leq \varepsilon \leq \varepsilon_{\max},$$

где $\varepsilon_{\min} > 0$, $R_L(\varepsilon_{\max}) = 0$ и $h(z) = -z \ln z - (1-z) \ln(1-z)$. Наибольшее значение $R_L(\varepsilon_{\min}) = I(\mathbf{X}, \Omega^c)$ достигается на байесовском тест-канале $\mathbf{Q}_{\varepsilon_{\min}} = \{P(\omega_j|\mathbf{x}) = P(\omega_j)p(\mathbf{x}|\omega_j) / \sum_{i=1}^c P(\omega_i)p(\mathbf{x}|\omega_i), j = 1, \dots, c\}$.

A data classification model based on the average mutual information

Mikhail Lange^{1*}

lange_mm@ccas.ru

*Andrey Lange*¹

lange_am@mail.ru

¹Moscow, FRC CSC RAS

Using the known source coding model with a given fidelity for the symbols transmitted by a noisy channel, an information theoretical classification model is suggested. Let $\Omega^c = \{\omega_1, \dots, \omega_c\}$, $c \geq 2$, be a set of classes (source alphabet) of prior probabilities $P(\omega_i) > 0$, $i = 1, \dots, c$ and \mathbf{X} be a set of the channel outputs (objects) with the class-conditional probability densities $\{p(\mathbf{x}|\omega_i), i = 1, \dots, c\}$, $\mathbf{x} \in \mathbf{X}$. A set of decisions $\Omega = \Omega^c$ on the objects from \mathbf{X} is made at an output of a test-channel $\mathbf{Q} = \{Q(\omega_j|\mathbf{x}), j = 1, \dots, c\}$, $\omega_j \in \Omega$, that is given by the probabilities of the class estimations on the object. The triple $\Omega^c, \mathbf{X}, \Omega$, together with the above distributions and densities produce the classification model

$$\Omega^c \Rightarrow \boxed{\{p(\mathbf{x}|\omega_i), i = 1, \dots, c\}} \Rightarrow \mathbf{X} \Rightarrow \boxed{\{Q(\omega_j|\mathbf{x}), j = 1, \dots, c\}} \Rightarrow \Omega$$

in which the average mutual information $I_{\mathbf{Q}}(\mathbf{X}; \Omega) \geq 0$ and the error rate $E_{\mathbf{Q}}(\mathbf{X}, \Omega) > 0$ are the functionals of the test-channel \mathbf{Q} . There is known the rate-distortion function as follows

$$R(\varepsilon) = \min_{\mathbf{Q}: E_{\mathbf{Q}}(\mathbf{X}, \Omega) \leq \varepsilon} I_{\mathbf{Q}}(\mathbf{X}; \Omega),$$

where the minimum is taken over \mathbf{Q} satisfying the error rate limitation by a given $\varepsilon > 0$. This function yields the minimal error rate ε subject to a fixed value of the average mutual information.

It has been obtained the lower bound $R_L(\varepsilon) \leq R(\varepsilon)$ as the decreasing function

$$R_L(\varepsilon) = I(\mathbf{X}; \Omega^c) - h(\varepsilon - \varepsilon_{\min}) - (\varepsilon - \varepsilon_{\min}) \ln(c-1), \quad \varepsilon_{\min} \leq \varepsilon \leq \varepsilon_{\max},$$

where $\varepsilon_{\min} > 0$, $R_L(\varepsilon_{\max}) = 0$ and $h(z) = -z \ln z - (1-z) \ln(1-z)$. The largest value $R_L(\varepsilon_{\min}) = I(\mathbf{X}, \Omega^c)$ is provided by the Bayes test-channel $\mathbf{Q}_{\varepsilon_{\min}} = \{P(\omega_j|\mathbf{x}) = P(\omega_j)p(\mathbf{x}|\omega_j) / \sum_{i=1}^c P(\omega_i)p(\mathbf{x}|\omega_i), j = 1, \dots, c\}$.

Получено обобщение для модели классификации с опцией отказа, в которой множество классов $\Omega^{c+1} = \Omega^c \cup \omega_0$, $c \geq 1$, с неизвестными априорными вероятностями $\theta = \sum_{i=1}^c P(\omega_i)$ и $1 - \theta = P(\omega_0)$ включает класс $\omega_0 \notin \Omega^c$ для реализации отказа. При $\theta \rightarrow 1$, для модели с отказом построена функция $\tilde{R}(\varepsilon_\theta) = \theta R(\varepsilon)$ от вероятности ошибки $\varepsilon_\theta = \theta^2 \varepsilon + 2\theta(1 - \theta) \geq \varepsilon$ и найдена ее нижняя граница $\tilde{R}_L(\varepsilon_\theta) = \theta R_L(\varepsilon)$ в диапазоне $\varepsilon_{\theta_min} \leq \varepsilon_\theta \leq \varepsilon_{\theta_max}$. В случае $\theta \leq 1$ уменьшение вероятности ошибки может быть достигнуто на ансамбле источников $\mathbf{X}^M = (\mathbf{X}_1, \dots, \mathbf{X}_M)$ размера $M \geq 2$, для которого $I(\mathbf{X}^M; \Omega^c) > I(\mathbf{X}_m; \Omega^c)$, $m = 1, \dots, M$.

Рассмотрены две схемы комплексирования данных источников: на основе голосования решений по источникам (Weighted Majority Vote) и на основе обобщенной меры различия объектов на ансамбле (General Dissimilarity Measure). Ниже даны экспериментальные оценки вероятностей ошибки распознавания лиц по декоррелированным компонентам $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$, RGB изображений и по ансамблю $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ с использованием схем комплексирования WMV и GDM. Оценки получены для NN и SVM классификаторов типа "класс против всех" без отказа ($\theta = 1$) и с отказом ($\theta = 0.5$). Экспериментальные результаты подтверждают повышение вероятности ошибки для модели с отказом по сравнению с моделью без отказа. Во всех случаях приведенные оценки демонстрируют уменьшение вероятности ошибки на ансамбле источников и преимущество схемы GDM по сравнению со схемой WMV. Полная версия работы представлена в [1].

| | θ | \mathbf{X}_1 | \mathbf{X}_2 | \mathbf{X}_3 | WMV | GDM |
|------------|----------|----------------|----------------|----------------|---------|---------|
| NN | 1 | 0,0042 | 0,0010 | 0,0016 | 0,00004 | 0,00002 |
| SVM | 1 | 0,0024 | 0,0027 | 0,0025 | 0,00020 | 0,00001 |
| NN | 0.5 | 0,0055 | 0,0145 | 0,0295 | 0,0035 | 0,0030 |
| SVM | 0.5 | 0,0190 | 0,0500 | 0,0540 | 0,0080 | 0,0025 |

Работа поддержана РФФИ, проект 18-07-01231.

- [1] Ланге М.М., Ланге А.М. О теоретико-информационной модели классификации данных // Машинное обучение и анализ данных, 2018.

For the model with a reject option, a generalization is developed. Such model contains a set of the classes $\Omega^{c+1} = \Omega^c \cup \omega_0$, $c \geq 1$ of unknown prior probabilities $\theta = \sum_{i=1}^c P(\omega_i)$ and $1 - \theta = P(\omega_0)$, where the zeroth class $\omega_0 \notin \Omega^c$ is used for making a reject decision. For $\theta \rightarrow 1$, the similar rate-distortion function $\tilde{R}(\varepsilon_\theta) = \theta R(\varepsilon)$ of the error rate $\varepsilon_\theta = \theta^2 \varepsilon + 2\theta(1 - \theta) \geq \varepsilon$ has been constructed and its lower bound $\tilde{R}_L(\varepsilon_\theta) = \theta R_L(\varepsilon)$ has been found in the segment $\varepsilon_{\theta_{\min}} \leq \varepsilon_\theta \leq \varepsilon_{\theta_{\max}}$. For $\theta \leq 1$ the error rate can be decreased in an ensemble of the data sources $\mathbf{X}^M = (\mathbf{X}_1, \dots, \mathbf{X}_M)$ of size $M \geq 2$ subject to the condition $I(\mathbf{X}^M; \Omega^c) > I(\mathbf{X}_m; \Omega^c)$, $m = 1, \dots, M$.

Two fusion schemes are investigated, namely the known WMV scheme by weighted majority voting the individual source decisions and by the original GDM scheme, which decision rule uses a general dissimilarity measure in the ensemble of the sources. The experimental face recognition error rates are shown for the decorrelated components \mathbf{X}_1 , \mathbf{X}_2 , \mathbf{X}_3 of RGB images and for the ensemble $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ using WMV and GDM fusion schemes. The results have been obtained by the multiclass NN and SVM classifiers of type "class-vs-all" without reject ($\theta = 1$) and with the reject option ($\theta = 0.5$). The experimental results support the higher error rates for the classifiers with reject as against the classifiers without reject. In all of the cases, the ensemble yields the less error rates in comparison with the individual sources. Also, GDM scheme demonstrates some advantage with respect to WMV scheme. The full version of the paper is prepared in [1].

| | θ | \mathbf{X}_1 | \mathbf{X}_2 | \mathbf{X}_3 | WMV | GDM |
|------------|----------|----------------|----------------|----------------|---------|---------|
| NN | 1 | 0,0042 | 0,0010 | 0,0016 | 0,00004 | 0,00002 |
| SVM | 1 | 0,0024 | 0,0027 | 0,0025 | 0,00020 | 0,00001 |
| NN | 0.5 | 0,0055 | 0,0145 | 0,0295 | 0,0035 | 0,0030 |
| SVM | 0.5 | 0,0190 | 0,0500 | 0,0540 | 0,0080 | 0,0025 |

The research is supported by RFBR project 18-07-01231.

- [1] *Lange M., Lange A.* On Information Theoretical Model for Data Classification // *Journal of Machine Learning and Data Analysis*, 2018.

Методы изучения локальных свойств функциональной структуры человеческого тела

Устинин Михаил Николаевич^{1*} u_m_n@mail.ru

*Рыкунов Станислав Дмитриевич*¹ stanislavrykunov@gmail.com

*Рыкунова Елена Дмитриевна*¹ alenarykunova@gmail.com

*Бойко Анна Ивановна*¹ a.boiko@list.ru

¹Пуццино, ИМПБ РАН - филиал ИПМ им. М.В. Келдыша РАН

DOI: 10.30826/IDP201871

Создан метод виртуальных электродов для изучения электрической активности головного мозга человека с помощью неинвазивных внешних измерений магнитного поля. По данным магнитной энцефалографии строится функциональная томограмма, которая отображает на пространство головы информацию, содержащуюся в многомерных временных рядах. Это достигается решением обратной задачи для всех элементарных осцилляций, выделяемых с помощью преобразования Фурье. Каждой частоте осцилляции соответствует узел трехмерной сетки, в котором располагается источник. Пользователь задает местоположение, размер и форму области мозга для детального изучения частотной структуры функциональной томограммы – виртуальный электрод. Совокупность осцилляций, попавших в заданную область, представляет собой парциальный спектр данной области. С помощью этого спектра можно восстановить временные ряды энцефалограммы, измеренной виртуальным электродом. Анализ корреляций между временными рядами, рассчитанными для разных точек мозга, позволит изучать его коннективность. Также можно оценить преимущественное направление источников, расположенных в объеме виртуального электрода.

Исследование выполнено за счет гранта Российского научного фонда (проект № 18-11-00178).

- [1] *Llinás R. R., Ustinin M. N., Rykunov S. D., Boyko A. I., Sychev V. V., Walton K. D., Rabello G. M., Garcia J.* Reconstruction of human brain spontaneous activity based on frequency-pattern analysis of magnetoencephalography data. // *Frontiers in Neuroscence*, 2015 9, 373. <http://doi.org/10.3389/fnins.2015.00373>

Methods to study local properties of the human body functional structure

*Ustinin Mikhail*¹*

u_m_n@mail.ru

*Rykunov Stanislav*¹

stanislavrykunov@gmail.com

*Rykunova Elena*¹

alenarykunova@gmail.com

*Boyko Anna*¹

a.boyko@list.ru

¹Pushchino, IMPB RAS - Branch of KIAM RAS

The method of virtual electrodes for studying electrical activity of the human brain using non-invasive external measurements of the magnetic field was developed. From the magnetic encephalography data, a functional tomogram is calculated, projecting onto the head space the information contained in multichannel time series. This is achieved by solving the inverse problem for all the elementary oscillations that are separated by the Fourier transform. For each oscillation frequency the node of a three-dimensional grid is found, where the source is located. User specifies the location, size and shape of the brain region for a detailed study of the frequency structure of the functional tomogram - a virtual electrode. Set of oscillations that fall in the given region is the partial spectrum of the given region. Using this spectrum, it is possible to reconstruct the time series of the encephalogram measured by the virtual electrode. Analysis of correlations between time series calculated for different points of the brain will make it possible to study its connectivity. Also it is possible to estimate the dominant direction of the sources inside the volume of the virtual electrode.

The research was supported by the Russian Science Foundation (grant 18-11-00178).

- [1] *Llinás R. R., Ustinin M. N., Rykunov S. D., Boyko A. I., Sychev V. V., Walton K. D., Rabello G. M., Garcia J.* Reconstruction of human brain spontaneous activity based on frequency-pattern analysis of magnetoencephalography data. // *Frontiers in Neuroscience*, 2015 9, 373. <http://doi.org/10.3389/fnins.2015.00373>

Исследование проблемы множественного сравнения при анализе всплескообразной электрической активности мозга и мышц

Сушкова Ольга Сергеевна^{1*} o.sushkova@mail.ru
*Морозов Алексей Александрович*¹ morozov@cplire.ru
*Габова Александра Васильевна*² agabova@yandex.ru
*Карabanов Алексей Вячеславович*³ doctor.karabanov@mail.ru

¹Москва, ИРЭ им. В.А. Котельникова РАН

²Москва, ИВНД и НФ РАН

³Москва, ФГБНУ «Научный центр неврологии»

DOI: 10.30826/IDP201872

Разработан новый метод анализа всплескообразной электрической активности мозга и мышц, основанный на вейвлет-анализе и ROC-анализе, который позволяет выявлять частотно-временные особенности сигналов электроэнцефалограмм (ЭЭГ), электромиограмм (ЭМГ) и акселерометра (АКС) пациентов с болезнью Паркинсона (БП). Идея метода заключается в поиске локальных максимумов на вейвлет-спектрограмме и вычисления различных характеристик, описывающих эти максимумы (всплески). Метод предусматривает перебор большого количества диапазонов выбранных характеристик, поэтому при статистическом оценивании различных гипотез возникает проблема множественного сравнения. Необходимо найти компромисс между степенью детализации изучаемых характеристик и величиной поправки Бонферрони. В докладе рассказывается о применении статистического оценивания статистических гипотез на данных пациентов с ранним паркинсонизмом.

Работа выполнена за счёт средств государственного задания № 0030-2015-0189, стипендии Президента РФ молодым учёным и аспирантам № СП-5247.2018.4 и поддержана Российской академией наук.

- [1] *Сушкова О. С., Морозов А. А., Габова А. В., Карabanов А. В.* Применение метода анализа всплескообразной электрической активности мозга для выявления особенностей ЭЭГ пациентов на ранней стадии болезни Паркинсона // Журнал неврологии и психиатрии им. С.С. Корсакова. — 2018. — Т. 118, № 7. — С. 45-48. — DOI: 10.17116/jnevto20181187145.

Investigation of the multiple comparisons problem in the wave train electrical activity analysis of the brain and limbs

Olga Sushkova^{1*}

`o.sushkova@mail.ru`

*Alexei Morozov*¹

`morozov@cplire.ru`

*Alexandra Gabova*²

`agabova@yandex.ru`

*Alexei Karabanov*³

`doctor.karabanov@mail.ru`

¹Moscow, Kotelnikov IRE RAS

²Moscow, IHNA & NF RAS

³Moscow, FSBI “Research Center of Neurology”

We have developed a new method for analyzing wave train electrical activity of the brain and muscles based on wavelet analysis and ROC analysis that enables to study the time-frequency features of electroencephalograms (EEG), electromyograms (EMG), and acceleration (ACC) signals in patients with Parkinson’s disease (PD). The idea of the method is to find local maxima in the wavelet spectrogram and to calculate various characteristics describing these maxima (named wave trains). The method involves a search of a large number of ranges of selected characteristics; therefore a problem of multiple comparison appears during statistical estimation of different hypotheses. It is necessary to find a compromise between the degree of detail of the investigated characteristics and the magnitude of the Bonferroni correction. The report describes the use of statistical estimation of statistical hypotheses on the data of patients with early Parkinson’s disease.

This research is funded by the Ministry of Science and Higher Education of Russian Federation, project No. 0030-2015-0189; the Scholarship of the President of the Russian Federation to Young Scientists and Post-graduate Students, grant No. CII-5247.2018.4; and Russian Academy of Sciences.

- [1] *Sushkova O. S., Morozov A. A., Gabova A. V., Karabanov A. V.* Application of brain electrical activity burst analysis method for detection of EEG characteristics in the early stage of Parkinson’s disease // *Zhurnal nevrologii i psikiatrii imeni S.S. Korsakova*. — 2018. — V. 118, No. 7. — P. 45-48. — DOI: 10.17116/jnevro20181187145.

Основанная на тестовых методах распознавания образов прикладная интеллектуальная система определения состояния пробуждения пациента во время общей анестезии

Янковская Анна Ефимовна^{1,2*} ayuankov@gmail.com

Пеккер Яков Семенович^{3,4} pekker@ssmu.ru

Бразовский Константин Станиславович^{3,4} bks_2005@mail.ru

¹Россия, Томск, ТГАСУ

²Россия, Томск, НИ ТГУ

³Россия, Томск, СибГМУ

⁴Россия, Томск, НИ ТПУ

DOI: 10.30826/IDP201873

Пробуждение пациента во время общей анестезии является серьезной клинической проблемой. Практика, направленная на снижение риска пробуждения, заключается в непрерывном внутриоперационном нейромониторинге. К сожалению, в некоторых случаях применяемые методы обработки электроэнцефалограмм не могут предсказать наступление пробуждения и надежно его детектировать. Предлагается подход к определению состояния пробуждения во время наркоза, основанный на тестовых методах распознавания образов и реализованный в прикладной интеллектуальной системе определения состояния пробуждения пациента (ИС ОСПП) во время общей анестезии. ИС ОСПП конструируется на базе интеллектуального инструментального средства ИМС-ЛОГ, основанном на матричном представлении данных и знаний и предназначенном для выявления закономерностей и принятия диагностических, организационно-управленческих и классификационных решений, а также обоснования принимаемых решений с использованием средств когнитивной графики. Предложенный подход сигнализирует с высокой степенью чувствительности и специфичности о возможном пробуждении пациента. Работа выполнена при поддержке гранта РФФИ № 16-07-00859.

- [1] *Янковская А. Е., Пеккер Я. С., Бразовский К. С.* Основанная на тестовых методах распознавания образов прикладная интеллектуальная система определения состояния пробуждения пациента во время общей анестезии // Машинное обучение и анализ данных, 2018 (в печати).

Intelligent system based on test methods of pattern recognition to detect awareness state during general anesthesia

Yankovskaya Anna^{1,2*}

ayyankov@gmail.com

Pekker Jakob^{3,4}

pekker@ssmu.ru

Brazovskii Konstantin^{3,4}

bks_2005@mail.ru

¹Russia, Tomsk, TSUAB

²Russia, Tomsk, NR TSU

³Russia, Tomsk, SSMU

⁴Russia, Tomsk, NR TPU

Awareness during general anesthesia, when a patient undergoes surgery, is quite a serious clinical problem. Frequently, complications manifest itself as a severe pain and post-traumatic stress disorder. Common practice to reduce the risk of developing such a condition is to monitor electrical activity of the patient's brain. Unfortunately, in some cases conventional methods to process electroencephalography (EEG) data fail to predict the upcoming state of awareness and can not indicate the awareness when it appears. A new approach to detect state of awareness based on test methods of pattern recognition is proposed and realized as an Intelligent Awareness State Detection System (IA SDS) intended to use during general anesthesia. The IA SDS is constructed over IMSLOG, an intelligent software tool realizing a matrix model to represent both data and knowledge. This software allows data mining and exploring variety of interconnections as well as making diagnostics, management and classification decisions. The software is also capable to substantiate the decisions using cognitive graphics.

The proposed approach can be used to detect the awareness state with both high sensitivity and specificity.

The research is supported by RFBR grant # 16-07-00859.

- [1] *Yankovskaya A.E., Pekker J.S., Brazovskii K.S.* Intelligent system based on test methods of pattern recognition to detect awareness state during general anesthesia // Machine Learning and Data Analysis, 2018 (in press).

Конвергенция нескольких методов для интеллектуального анализа данных и знаний

Анна Ефимовна Янковская^{1,2,*}

ayyankov@gmail.com

*Ольга Михайловна Гергет*³

olgagerget@mail.ru

¹Россия, Томск, ТГАСУ

²Россия, Томск, НИ ТГУ

³Россия, Томск, НИ ТПУ

DOI: 10.30826/IDP201874

Применение конвергенции нескольких наук и научных направлений при решении задач интеллектуального анализа данных и знаний не вызывает сомнений. Дополнительно к применяемым методам в конвергенцию включен метод машинного обучения, основанный на нейросетевом подходе. Для интеллектуального анализа динамических данных применяют глубокие нейронные сети. Проблемой обучения таких сетей является затухание локального градиента. Частично решает данную проблему использование алгоритма Resilient propagation. При обучении сетей алгоритм Resilient propagation использует коррекцию веса, основанную на динамике знака производных к совместно применяемым весам t раз, где t – это количество развертываний нейросети. При этом наблюдается зависимость коррекции веса от временного интервала, на котором получены исходные данные. В модификации алгоритма обучения применяется коррекция веса в зависимости от динамики знака суммы производных целевой функции по совместно используемому весовому коэффициенту, т.е. коррекция применяется один раз, что позволяет исключить проблему зависимости от длины рассматриваемого временного интервала и чрезмерно высокой коррекции веса. Ранее реализованные методы лежат в основе интеллектуального инструментального средства ИМСЛОГ. В дальнейшем будет осуществлён анализ результатов исследования с применением предлагаемой конвергенции методов с учётом их весовых коэффициентов. [1].

Работа поддержана грантом РФФИ №16-07-00859 и ГЗ «Наука» № № 17.8205.2017

- [1] *Янковская А. Е., Гергет О. М.* Конвергенция нескольких методов для интеллектуального анализа данных и знаний // Машинное обучение и анализ данных, 2018 (в печати).

Convergence of several methods for intelligent analysis of data and knowledge

Yankovskaya Anna^{1,2,*}

ayyankov@gmail.com

*Gerget Olga*³

olgagerget@mail.ru

¹City, Institution

²City, Institution

Application of convergence of several sciences and scientific directions in solving problems of data mining and knowledge is beyond doubt. In addition to the methods used earlier this publication uses convergence with machine learning method based on the neural network approach. Deep neural networks are now being used very successfully for analysis dynamic data. However, the well-known problem of learning networks is the attenuation of the local gradient. The Resilient propagation algorithm partly solve this problem. When learning deep neural networks, the Resilient propagation algorithm uses a weight correction based on the dynamics of the sign of the derivatives to the jointly used weights t - times, if t is the number of neural network deployments. In this case, it is the observed dependence of the weight correction on the time interval. The proposed modification of the learning algorithm uses weight correction depending on the sign dynamics of the sum of derivatives of the goal function by the jointly used weight coefficient. Thus, the correction is applied once to use in combination with weight coefficient, that solves the problem of dependence the function on the length of the time interval and excessively high weight correction. Previously implemented methods are the basis of the intelligent integration software IMSLOG. In the future, it is proposed to analyze the results of the research with making use of the convergence of methods with account to weight coefficients [1].

This research is funded by RFBR, grant # 16-07-00859 and State Assignment "Science" No. # 17.8205.2017.

- [1] *Yankovskaya A. E., Gerget O. M.* Convergence of several methods for intelligent analysis of data and knowledge // Machine learning Data Analysis, 2018 (in press).

Обнаружение отклонений на маммограммах с использованием интерфейса мозг-компьютер

*Сулимова Валентина Вячеславовна** vsulimova@yandex.ru

Россия, Тула, ТулГУ

Красоткина Ольга Вячеславовна O.V.krasotrina@yandex.ru

Бухонов Сергей Александрович king1677@mail.ru

Russia, Moscow, МИРТ

Моттль Вадим Вячеславович vmottl@yandex.ru

Russia, Moscow, Computing Center of the RAS

Дэвид Уиндридж D.Windridge@mdx.ac.uk

UK, London, Middlesex University

DOI: 10.30826/IDP201875

В последнее время появилась тенденция применения интерфейса мозг-компьютер для обнаружения разнообразных целевых объектов на изображениях. В частности, несколько лет назад нашими коллегами из Великобритании была предложена технология обнаружения отклонений на маммограммах, основанная на анализе мозговой активности эксперта в процессе просмотра быстро предъявляемых ему изображений маммограмм. Высокая пропускная способность данного подхода обусловлена быстротой работы человеческого мозга, способного за доли секунды узнать заданный объект.

В данной работе мы усовершенствуем указанную технологию за счет, во-первых, применения специальной предварительной обработки сигналов мозговой активности, поступающих с 66 электродов, закрепленных на голове испытуемых в процессе просмотра ими изображений маммограмм и, во-вторых, применения метода опорных векторов с двумя видами регуляризации, учитывающими особенности рассматриваемой задачи, что позволяет ощутимо повысить качество обнаружения маммограмм с отклонениями.

Работа выполнена при поддержке РФФИ, грант № 18-07-01087.

- [1] *Sulimova V. V., Krasotkina O. V., Buchonov S. A., Mottl V. V., Windridge D.* Regularized SVMs for classification of image evoked EEG potentials captured from an observer.

Detection of abnormalities in mammograms using the brain-computer interface

*Sulimova Valentina**

vsulimova@yandex.ru

Russia, Tula, Tula State University

Krasotkina Olga

O.V.krasotrina@yandex.ru

USA, Summit, Markov Processes International

Buchonov Sergey

king1677@mail.ru

Russia, Moscow, MIPT

Mottl Vadim

vmottl@yandex.ru

Russia, Moscow, Computing Center of the RAS

Windridge David

D.Windridge@mdx.ac.uk

UK, London, Middlesex University

Last few years a trend of using brain-computer interfaces (BCIs) for identifying targets within different image types has emerged. Particularly, some years ago our colleagues from Great Britain first proposed to use BCI as a novel high throughput screening technology for mammography. This technology is based on rapid serial visual presentation of mammograms images to experts and recognition of its brain activity that corresponds to seeing mammograms with abnormalities. High performance is provided by the advanced computing power of the human visual system, which has been found to be able to identify and classify the image content as target or non-target with a frequency of more than 10 Hz.

In this paper, we improve this technology by, firstly, applying special pre-processing of brain activity signals coming from 66 electrodes fixed on the head of the test persons while they are viewing mammogram images and, secondly, by applying the support vector machines method with two kinds of regularization, taking into account the peculiarities of the considered problem, which makes it possible to significantly improve the quality of detection of abnormal mammograms.

This research is funded by RFBR, grant 18-07-01087.

- [1] *Sulimova V. V., Krasotkina O. V., Buchonov S. A., Mottl V. V., Windridge D.* Regularized SVMs for classification of image evoked EEG potentials captured from an observer.

Анализ торсионных углов между осями спиралей в спиральных парах белковых молекул

Тихонов Дмитрий Анатольевич^{1,3} dmitry.tikhonov@gmail.com

Куликова Людмила Ивановна^{1*} likulikova@mail.ru

*Ефимов Александр Васильевич*² efimov@protres.ru

¹Россия, Пущино, ИМПБ РАН, филиал ИПМ им. М.В.Келдыша РАН

²Россия, Пущино, Институт белка РАН

³Россия, Пущино, Институт Теоретической и Экспериментальной
Биофизики РАН

DOI: 10.30826/IDP201876

В данной работе проведен анализ распределения торсионных углов Ω между осями спиралей в парах связанных между собой перетяжками спиралей в пространственных структурах белковых молекул. Исследование проводилось на множестве спиральных пар, отобранных из структур белковых молекул, представленных в PDB. Полученное множество спиральных пар было проанализировано и разбито на три подмножества по критерию пересечения проекций спиралей на параллельные плоскости, проходящие через оси спиралей. Показано, что спиральные пары, образованные двумя α -спиралями, представляют значительную часть подмножества с пересечением проекций спиралей, и распределение таких структур имеет максимум в области значений торсионного угла между осями спиралей в области $20^\circ < \Omega < 25^\circ$. Абсолютное большинство пар подмножества с пересечением осей спиралей составляют спиральные пары, образованные двумя α -спиралями. Для них характерно распределение с тремя максимумами, лежащими в области острых углов: один — в области отрицательных значений $-50^\circ < \Omega < -25^\circ$, другой — в области положительных значений углов $20^\circ < \Omega < 25^\circ$ и еще одним в области прямого угла $70^\circ < \Omega < 110^\circ$.

Работа поддержана грантами РФФИ № 16-01-00692 и № 18-07-01031.

- [1] *Тихонов Д. А., Куликова Л. И., Ефимов А. В.* Анализ торсионных углов между осями спиралей в спиральных парах белковых молекул // Математическая биология и биоинформатика, 2017. Т. 12. № 2. С. 398–410. doi: 10.17537/2017.12.398.

Analysis of Torsion Angles between Helical Axes in Pairs of Helices in Protein Molecules

Tikhonov Dmitrii^{1,3}

dmitry.tikhonov@gmail.com

Kulikova Liudmila^{1*}

likulikova@mail.ru

*Efimov Aleksandr*²

efimov@protres.ru

¹Pushchino, Russia, IMPB RAS — Branch of KIAM RAS

²Pushchino, Russia, Institute of Protein Research RAS

³Pushchino, Russia, Institute of Theoretical and Experimental Biophysics RAS

In this study, an analysis of distribution of the torsion angles Ω between helical axes in pairs of connected helices found in known proteins has been performed. The database for helical pairs was compiled using the Protein Data Bank taking into account the definite rules suggested earlier. The database was analyzed in order to elaborate its classification and find out novel structural features in helix packing. The database was subdivided into three subsets according to criterion of crossing helix projections on the parallel planes passing through the axes of the helices. It was shown that helical pairs not having crossing projections are distributed along the whole range of angles Ω , although there are two maxima at $\Omega = 0^\circ$ and $\Omega = 180^\circ$. Most of helical pairs of this subset are pairs formed by α -helices and 3_{10} -helices. It is shown that the distribution of all the helical pairs having the crossing helix projections has a maximum at $20^\circ < \Omega < 25^\circ$. In this subset, most helical pairs are formed by α -helices. The distribution of only α -helical pairs having crossing axes projections has three maxima, at $-50^\circ < \Omega < -25^\circ$, $20^\circ < \Omega < 25^\circ$, and $70^\circ < \Omega < 110^\circ$.

This research is funded by RFBR, grants 16-01-00692, 18-07-01031.

- [1] *Tikhonov, D., L. Kulikova, and A. Efimov.* Analysis of Torsion Angles between Helical Axes in Pairs of Helices in Protein Molecules // *Math. Biol. Bioinform.*, 2018. 13. Suppl. pp. t17–t28. doi: 10.17537/2018.13.t17.

Жестовое управление дополненной реальностью при кардиологических интервенциях

Васильев Евгений Павлович^{1*} evgeny.vasiliev@itmm.unn.ru

*Takashi Komuro*² komuro@mail.saitama-u.ac.jp

*Еурлапов Вадим Евгеньевич*¹ vadim.turlapov@gmail.com

*Никольский Александр Викторович*³ yahtingman@rambler.ru

¹Нижний Новгород, ННГУ им. Н.И. Лобачевского

²Сайтама, Университет Сайтама

³Нижний Новгород, Городская клиническая больница №5 г. Нижнего Новгорода

DOI: 10.30826/IDP201877

Работа интервенционного кардиолога аритмолога сопряжена с высокой концентрацией внимания в ходе операции по устранению аритмий. Врач одновременно контролирует показатели нескольких мониторов и индикаторных панелей приборов ЭФИ (электрофизиологической) лаборатории и навигационной системы (Carto).

Мы предлагаем интерфейс, который позволяет пользователю взаимодействовать с трехмерной моделью сердца, при помощи жестов, которые могут выполняться одной рукой без взаимодействия с мышью и другими физическими устройствами ввода, и таким образом перемещать, масштабировать и вращать модель.

Использование только одной руки для взаимодействия с моделью более удобно для человека, чем использование двух рук, а в определенных ситуациях является единственно возможным типом взаимодействия.

В ходе исследований было создано демонстрационное приложение, чтобы показать преимущества этого подхода.

Работа поддержана Министерством Образования и Науки Российской Федерации (Контракт № 02.G25.31.0157)

- [1] *Васильев Е. П., Комуро Т., Турлапов В. Е., Никольский А. В.* Жестовое управление дополненной реальностью при кардиологических интервенциях // Машинное обучение и анализ данных, Москва: Издательство, 2018. — С. 89–96. <http://jmla.org/papers/doc/2018/no2/Vasiliev2018OneHand.pdf>.

One hand aerial gesture control for AR-based cardiac interventions

Evgeny Vasiliev^{1*}

evgeny.vasiliev@itmm.unn.ru

*Takashi Komuro*²

komuro@mail.saitama-u.ac.jp

*Vadim Turlapov*¹

vadim.turlapov@gmail.com

*Alexander Nikolsky*³

yahtingman@rambler.ru

¹Nizhny Novgorod, Lobachevsky University

²Saitama, Saitama University

³Nizhny Novgorod, Municipal Clinical Hospital 5 of Nizhny Novgorod

We consider the problem of interaction of the operating surgeon with medical software during the operations.

We propose an interface that allows the user to interact with a cg model using one hand gestures, executing operations that can be performed with one hand are moving, zooming and rotating.

Using only one hand to interact with the model is more convenient for a person than using two hands.

We implemented three type of cursor: Without pointer (This option is suitable for touch panels, because in this case the user's finger acts as pointer), simple dot (Easy to create, but not very descriptive, especially when moving in three dimensions), virtual hand (Very descriptive in 3D space, but very difficult to create, off the shelf implementation is not available)

We created a demo application to show the advantages of this approach.

This research is funded by Ministry of Education and Science of Russia (Contract 02.G25.31.0157).

- [1] *Vasiliev E., Komuro T., Turlapov V., Nikolsky A.* One hand aerial gesture control for AR-based cardiac interventions // Machine Learning and Data Analysis, Moscow: Publisher, 2018. — p.89–96. <http://jmla.org/papers/doc/2018/no2/Vasiliev2018OneHand.pdf>.

Разработка системы для моделирования рельефа дна Северного Ледовитого океана

| | |
|---|------------------------|
| <i>Флоринский Игорь Васильевич</i> ¹ ★ | iflor@mail.ru |
| <i>Филитов Сергей Валерьевич</i> ¹ | fsv141@mail.ru |
| <i>Абрамова Анастасия Сергеевна</i> ² | abramanastas@gmail.com |
| <i>Зарайская Юлия Андреевна</i> ² | geozar@yandex.ru |
| <i>Селезнева Евгения Викторовна</i> ¹ | seevgenia@gmail.com |

¹Пуццо, ИМПБ РАН — филиал ИПМ им. М.В. Келдыша РАН

²Москва, ГИН РАН

DOI: 10.30826/IDP201878

Нами разрабатывается информационно-вычислительная система для морфометрического моделирования рельефа дна Северного Ледовитого океана [1]. Разработка включает следующие этапы: Адаптация вычислительных алгоритмов геоморфометрии и интерактивной 3D визуализации к работе с большими цифровыми моделями рельефа (ЦМР); Создание программного обеспечения, реализующего указанные алгоритмы; Выбор наиболее точной открытой ЦМР дна океана для последующих расчетов и моделирования; Фильтрация выбранной ЦМР; Расчет моделей морфометрических характеристик по ЦМР для нескольких масштабных уровней; Создание интерактивных многомасштабных морфометрических 3D моделей рельефа дна; Создание геопортала для хранения моделей и их интерактивной многомасштабной 3D визуализации. Система обеспечит: а) хранение большой ЦМР дна океана; б) расчет на ее основе моделей морфометрических характеристик; в) интерактивную 3D многомасштабную визуализацию моделей; г) свободный доступ к этой информации через Интернет. Система создается для поддержки гидрографических, морских геоморфологических, геологических, геофизических и океанологических исследований Арктического региона. Работа поддерживается на грантом РФФИ № 18-07-00223.

- [1] *Florinsky I. V., Filippov S. V., Abramova A. S., Zarayskaya Yu. A., Selezneva E. V.* Towards geomorphometric modelling of the topography of the Arctic Ocean floor // *Bandrova T., Konečný M. (Eds.), Proceedings of the 7th International Conference on Cartography and GIS, June 18–23 2018, Sozopol, Bulgaria, Vol. 1.* — Sofia: Bulgarian Cartographic Association, 2018, pp. 166–173.

Development of a system for modelling the Arctic Ocean floor topography

Florinsky Igor^{1*}

iflor@mail.ru

*Filippov Sergei*¹

fsv141@mail.ru

*Abramova Anastasia*²

abramanastas@gmail.com

*Zarayskaya Yulia*²

geozar@yandex.ru

*Selezneva Evgenia*¹

seevgenia@gmail.com

¹Pushchino, IMPB KIAM RAS ²Moscow, GIN RAS

We develop an information computing system for geomorphometric modelling of the topography of the Arctic Ocean floor [1]. The system development includes the following main stages: (1) Adaptation of previously developed computational methods of geomorphometry and algorithms of interactive three-dimensional (3D) visualization to handle and process big digital elevation models (DEMs); (2) Development of software implementing these methods and algorithms; (3) Selection of the most accurate, free, publicly available DEM describing the topography of the Arctic Ocean floor; (4) Filtering (noise and artefact reduction) of the selected DEM; (5) Derivation of models of morphometric variables from the filtered DEM for several scales; (6) Creation of interactive multiscale 3D morphometric models; (7) Development of a geportal.

The system will provide: (a) Storage of a big DEM of the ocean floor; (b) Derivation of models of morphometric variables from this DEM; (c) Interactive multiscale 3D visualization of the obtained models; and (d) Free access to this information via Internet.

The system is created to support hydrographic, marine geomorphological, geological, geophysical, and oceanological studies of the Arctic Region.

This research is funded by RFBR, grant 18-07-00223.

- [1] Florinsky, I. V., S. V. Filippov, A. S. Abramova, Yu. A. Zarayskaya, and E. V. Selezneva. 2018. Towards geomorphometric modelling of the topography of the Arctic Ocean floor. In: *Bandrova, T., and Konečný, M. (Eds.), Proceedings of the 7th International Conference on Cartography and GIS, June 18-23 2018, Sozopol, Bulgaria, Vol. 1*. Bulgarian Cartographic Association, Sofia, pp. 166–173.

Синтез кривых ГИС с использованием генеративных конкурирующих глубоких конволюционных нейронных сетей

*Визильтер Юрий Валентинович*¹,

vis@gosniias.ru

Горбачев Владимир Сергеевич^{1*}

gvs@gosniias.ru

*Хайдаров Артур Салаватович*²

Khaydarov.AS@gazpromneft-ntc.ru

*Яковлев Владимир Александрович*²

Yakovlev.AAle@gazpromneft-ntc.ru

¹г. Москва, ФГУП ГосНИИАС

²г. Санкт-Петербург, Научно-Технический Центр «Газпром нефти»

DOI: 10.30826/IDP201879

Изучение геологического разреза и выявление полезных ископаемых в скважинах проводятся с использованием различных видов геофизических исследований. По анализу форм, амплитуд и виду каротажных диаграмм (сигналов от датчиков различной природы) принимается решение о последующем (более детальном) рассмотрении обнаруженных интервалов командой экспертов (петрофизиков, сейсмиков, геологов, разработчиков и т.д.). Существует ряд специфических проблем, значительно усложняющих работу геолога. Одной из них является частичное или полное отсутствие данных по некоторым физическим исследованиям для ряда скважин месторождения. В работе рассматривается проблема восстановления отсутствующего участка ГИС (или кривой целиком) путем использования скрытых закономерностей между различными геофизическими исследованиями скважин. Для решения данной задачи предлагается использование ГКНС оригинальной архитектуры. Для повышения качества восстановления применялся способ обучения генеративных конкурирующих ГКНС. Результаты тестирования показали, что в рамках этого объекта предлагаемый алгоритм позволяет восстанавливать ГИС с качеством, достаточным для последующего анализа экспертом-геологом.

- [1] *Визильтер Ю. В. Горбачев В. С. Хайдаров А. С. Яковлев А. А.* Синтез кривых ГИС с использованием генеративных конкурирующих глубоких конволюционных нейронных сетей // Вестник компьютерных и информационных технологий, 2018 (в печати).

Well LOG synthesis using generative adversarial CNN

*Vizilter Yury*¹

vis@gosniias.ru

Gorbatsevich Vladimir^{2*}

gvs@gosniias.ru

*Khaydarov Arthur*²

Khaydarov.AS@gazpromneft-ntc.ru

*Yakovlev Andrei*²

Yakovlev.AAle@gazpromneft-ntc.ru

¹Moscow, FGUP GosNIIAS

²Saint-Petersburg, Gazpromneft Science & Technology Centre

The practice of well logging in order to identify minerals in the wells is carried out using various types of geophysical techniques. After performing the primary analysis of the shapes, amplitudes and types of geophysical logs (signals from sensors of different nature), the decision is made about the necessity of subsequent (and more detailed) examination of the detected intervals with the involvement of experts (petrophysicists, seismologists, geologists, developers, etc.). There are a number of specific problems that can significantly complicate a geologist's work. Among them is the partial or total lack of well logging records for a number of wells in the field. The paper addresses the problem of restoration (synthesis) of missing parts of well logging curve (or the whole curve) by retrieving hidden data patterns in different well logging. To solve this task, the special architecture of a deep convolutional neural network (CNN) is proposed. To improve the restoration quality, the generative adversarial network (GAN) approach is used for CNN training. The obtained experimental results show that for the considered task the proposed algorithm allows a well logging record to be restored with a quality that is sufficient for its subsequent analysis by an expert geologist.

- [1] *Vizilter Y. Gorbatsevich V. Khaydarov A. Yakovlev A.* Well log synthesis using generative adversarial CNN // *Vestnik komp'yuternykh i informatsionnykh tekhnologii* (Herald of computer and information technologies), 2018 (in press).

Многофакторный метод геоэкологоохранного прогнозирования

Хайретдинов Марат Саматович^{1,3} marat@opg.sssc.ru

*Агафонов Вадим Михайлович*² agvadam@yandex.ru

*Ковалевский Валерий Викторович*¹ kovalevsky@sscc.ru

*Воскобойникова Гюльнара Маратовна*¹ gulya@opg.sssc.ru

¹Новосибирск, Институт вычислительной математики и математической геофизики СО РАН

²Москва, Московский физико-технический институт

³Новосибирск, Новосибирский государственный технический университет

DOI: 10.30826/IDP201880

Обосновывается и излагается предложенный авторами оригинальный инструментальный метод, направленный на предупреждение геоэкологических рисков для социальной и природной среды от мощных техногенных и природных взрывов. Известно, что последние являются источниками ударных воздушных и сейсмических волн, оказывающих разрушительное воздействие на промышленные и жилые объекты. При этом комплексное воздействие внешних метеофакторов (направление и скорость ветра, температура и влажность окружающего воздуха др.) на распространение акустических волн может приводить к многократному усилению разрушительного воздействия инфразвука от взрывов на окружающую среду на определенных азимутальных направлениях. Это определяет необходимость прогнозирования направления главного ударного воздействия с учетом метеофакторов. Авторами предлагается и анализируется подход к решению проблемы с применением вибрационных источников. Такой подход учитывает в себе совокупность факторов, обуславливающих повышенные метеозависимые риски. Значение рисков оценивается удельной плотностью акустической энергии $E = \frac{1}{\rho c} \int p^2(t) dt$. Здесь ρc – удельное акустическое сопротивление воздуха, равное 42 г/(см²·с); $p(t)$ – регистрируемое акустическое давление, T – длительность акустической волны. Допустимые акустические воздействия на объекты социальной инфраструктуры и человека определяются табулированными значениями удельной плотности энергии в единицах Дж/м².

Multifactorial method of geocological prediction

Khairtdinov Marat^{1,3}

marat@opg.sccc.ru

*Agaphonov Vadim*²

agvadim@yandex.ru

*Kovalevsky Valery*¹

kovalevsky@sscc.ru

*Voskoboynikova Gyulnara*¹

gulya@opg.sccc.ru

¹Novosibirsk, Institute of Computational Mathematics and Mathematical Geophysics SB RAS

²Moscow, Moscow Institute of Physics and Technology

³Novosibirsk, Novosibirsk State Technical University

The report presents and substantiates theoretically and experimentally the original instrumental method proposed by the authors aimed at preventing geocological risks for the social and natural environment from powerful technogenic and natural explosions. It is known that they are sources of shock air and seismic waves, which have a destructive effect on industrial and residential objects. In this case, the complex effect of external meteorological factors (direction and speed of wind, temperature and humidity of the surrounding air, etc.) on the propagation of acoustic waves can lead to a multiple increase in the destructive effect of infrasound from explosions on the environment on certain azimuthal directions. This determines the need to predict the direction of the main impact taking into account meteorological factors. The authors suggest and analyze the approach to solving the problem with the use of vibration sources. This approach takes into account a set of factors that cause increased meteorological risks. The significance of risks is estimated by the specific density of acoustic energy $E = \frac{1}{\rho c} \int p^2(t) dt$. Here ρc is the specific acoustic resistance of air, equal to 42 g / (cm² · s); $p(t)$ is the registered acoustic pressure, T is the duration of the acoustic wave. Admissible acoustic effects on social infrastructure and human objects are determined by tabulated values of the specific energy density in J/m² units. Acoustic pressure is a function of many parameters determined by the radiation conditions and the long-range propagation of acoustic oscillations. The multifactorial model of integral pressure as the equation of energy balance can be represented in the form:

$$P_{\Sigma}(t, f, r) = P_s(f) + P_{abs} + P_{meteo}(e, \tau, \omega, \varphi) + P_{spher} + P_{sur}$$

Акустическое давление является функцией многих параметров, определяемых условиями излучения и дальнего распространения акустических колебаний. Многофакторная модель интегрального давления как уравнение энергетического баланса может представляется в виде:

$$P_{\Sigma}(t, f, r) = P_s(f) + P_{abs} + P_{meteo}(e, \tau, \omega, \varphi) + P_{spher} + P_{sur}$$

Здесь $P_{\Sigma}(t, f, r)$ – давление в точке регистрации на удалении r от источника; $P_s(f)$ – частотно зависимое акустическое давление от вибратора; P_{abs} – поглощение инфразвука по расстоянию; $P_{meteo}(e, \tau, \omega, \varphi)$ – давление, как функция метеопараметров: относительной влажности, температуры, направления и скорости ветра; φ – угол между направлением ветра и волновым фронтом от источника; P_{spher} – сферическая расходимость волнового фронта; P_{sur} – вариации давления из-за поглощения дневной поверхностью Земли.

Получение оценок рисков E в аналитическом виде в большинстве случаев наталкивается на сложности из-за отсутствия необходимой полноты априорных сведений, касающихся не только метеопараметров, но и сведений о пересеченности поверхности Земли и ее покрове (растительный и снежный покровы) и др. факторы. Получение аналитической зависимости возможно для некоторых частных случаев. Другой путь преодоления априорной неопределенности связан с изучением закономерности зависимости интегрального давления на основе экспериментов с применением вибраторов, а также взрывов в качестве излучателей инфранизкочастотных акустических колебаний. При этом учитывается вся полнота факторов, определяющих условия излучения и распространения в атмосфере акустических колебаний. Оба пути решения задачи рассматриваются в работе. Работа поддержана грантом РФФИ № 17-07-00872, 16-07-01052.

- [1] *M. Khairtdinov, V. Gubarev, et al.* Prediction of Environmental Risks from Explosions Based on a Set of Coupled Geophysical Field// Optoelectronics, Instrumentation and Data Processing, New York: Allerton Press, INC, 2014, Vol. 50, No. 4. P.3-13. Русский вариант в журнале "Автоматрия". 2014, Т.50, №4, с.с. 3-23 .

Here $P_{\Sigma}(t, f, r)$ is the pressure at the point of registration at a distance r from the source; $P_s(f)$ is frequency-dependent acoustic pressure from the vibrator; P_{abs} is absorption of infrasound by distance; $P_{meteo}(e, \tau, \omega, \varphi)$ is pressure, as a function of meteorological parameters: relative humidity, temperature, direction and wind speed; φ is the angle between the direction of the wind and the wave front from the source; P_{spher} is spherical divergence of the wave front; P_{sur} is pressure variations due to absorption by the Earth's surface.

The obtaining of risk assessments of E in analytical form in most cases encounters difficulties due to the lack of necessary completeness of a priori information concerning not only meteorological parameters, but also information on the intersection of the Earth's surface and its cover (vegetation and snow cover) and other factors. Obtaining analytic dependence is possible for some particular cases. Another way to overcome a priori uncertainty is to study the regularity of the dependence of integral pressure on the basis of experiments using vibrators, as well as explosions as emitters of infralow-frequency acoustic oscillations. In this case, all the factors that determine the conditions for the emission and propagation of acoustic vibrations in the atmosphere are taken into account. Both ways of solving the problem are considered in the paper.

This research is funded by RFBR, grants 17-07-00872 and 16-07-01052.

- [1] *M.S. Khairtdinov, V. V. Gubarev, V. V. Kovalevskii, S. A. Avrorov, G. M. Voskoboynikova, G. F. Sedukhina, A. A. Yakimenko* Prediction of Environmental Risks from Explosions Based on a Set of Coupled Geophysical Field// Optoelectronics, Instrumentation and Data Processing, New York: Allerton Press, INC, 2014, Vol. 50, No. 4. P.3-13. (Scopus).

Анализ и прогнозирование гидрологического ряда на основе обобщённых прецедентов

*Наумов Владимир Аркадьевич*¹ van-old@mail.ru

*Нелубина Елена Андреевна*¹ e.nelubina@gmail.com

*Рязанов Владимир Васильевич*² rvvccas@mail.ru

Виноградов Александр Петрович^{2*} vngrccas@mail.ru

¹Калининградский государственный технический университет,
Калининград, Россия

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

DOI: 10.30826/IDP201881

В работе представлен новый подход к использованию аппарата обобщённых прецедентов в задачах анализа и прогнозирования гидрологических рядов. Обобщённые прецеденты представляют собой вычислительный инструментарий, позволяющий на унифицированной основе задействовать априорные, непосредственно наблюдаемые или предпочтительные по тем или иным причинам локальные закономерности в данных. Представлены основные этапы схемы применения обобщённых прецедентов, показана тесная связь со схемой преобразования Хафа. Исследуются возможности сопоставления и совместного анализа метеоданных и фактических данных по объёму речного стока. В этом случае обобщённые прецеденты представляют собой нелинейные зависимости между определенными параметрами. Целью является выявление дифференциации регионов бассейна реки по их аккумулирующим возможностям. Рассматривается метод на основе анализе ограниченной по времени современной статистики. Полученные характеристики стока в регионах могут далее использоваться для краткосрочного прогноза вариаций уровня реки и других гидрологических процессов и явлений, в том числе паводков и тому подобных ситуаций. Эти характеристики служат важным фактором при исследовании экосистем, геологии региона и других целей. Работа поддержана грантом РФФИ.

- [1] *Наумов В. А., Нелубина Е. А., Рязанов В. В., Виноградов А. П.* Анализ и прогнозирование гидрологического ряда на основе обобщённых прецедентов // Машинное обучение и анализ данных, Москва: Вычислительный центр им. А.А. Дородницына РАН, 2018.

Analysis and prediction of hydrological series based on generalized precedents

*Naumov Vladimir Arkadievich*¹

van-old@mail.ru

*Nelyubina Elena Andreevna*¹

e.nelubina@gmail.com

*Ryazanov Vladimir Vassilievich*²

rvvccas@mail.ru

Vinogradov Alexander Petrovich^{2*}

vngrccas@mail.ru

¹Kaliningrad State Technical University, Kaliningrad, Russia

²Dorodnicyn Computing Centre of FRC CSC RAS, Moscow, Russia

The paper presents a new approach to the use of the apparatus of generalized precedents in the problems of analysis and prediction of hydrological series. Generalized precedents are computational tools that yield to use various a priori, directly observed or preferred for some or another reason local regularities in the data on a unified basis. The main stages of the scheme for applying generalized precedents are presented, and a close relationship is shown with the Hough transform scheme. The possibilities of comparison and joint analysis of meteorological data and actual data on the volume of river flow are investigated. In this case, the generalized precedents are typical nonlinear relationships between certain hydrological parameters. The goal is to identify the differentiation of the regions of the river basin by their accumulating capabilities. We show how this can be done on the basis of an analysis of time-limited contemporary statistics. Obtained flow characteristics in the regions can be further used for short-term forecasting of river level variations and other hydrological processes and phenomena, including flood and drought situations. These characteristics can also serve as an important factor in the study of ecosystems, geology of the region and other similar purposes.

This research is funded by RFBR.

- [1] *Naumov V., Nelyubina E., Ryazanov V., Vinogradov A.* Analysis and prediction of hydrological series based on generalized precedents // Machine Learning and Data Analysis, Moscow: Computing Centre of A.A. Dorodnicyn RAS, 2018.

Идентификация параметров модели

*Засухин Сергей Владимирович*¹

s.zasukhin@yandex.ru

Засухина Елена Семеновна^{2*}

elenazet2017@yandex.ru

¹Долгопрудный, Московский физико-технический институт

²Москва, ФИЦ «Информатика и управление» РАН

DOI: 10.30826/IDP201882

Задание точных значений гидрофизических почвенных параметров в гидрологических моделях имеет важнейшее значение при моделировании водного потока и переноса растворенных веществ в зоне аэрации, оценивании опасности катастрофических наводнений. Их экспериментальное определение является трудной задачей. Альтернативными подходами являются применение педотрансферных функций Rosetta, получаемых методом нейронных сетей, и обратное моделирование. Первый требует большого количества входных данных. Определение параметров в рамках второго подхода осуществляется, как правило, на основе наблюдаемых кривой водоудерживания и гидравлической проводимости. В настоящей работе рассматривается модель передвижения воды в ненасыщенной почве, в которой движение воды описывается уравнением Ричардса. Задача нахождения параметров формулируется как задача обратного моделирования, в которой искомые параметры определяются в результате подгонки моделируемых значений почвенной влажности к их наблюдаемым значениям. Численное решение задачи осуществляется методом наискорейшего спуска. При этом точные значения градиента целевой функции вычисляются по формулам быстрого автоматического дифференцирования. Численные расчеты показывают, что предложенный метод позволяет находить параметры с хорошей точностью. Особенности метода заключаются в применении быстрого автоматического дифференцирования в процессе оптимизации и в том, что параметры находятся в результате подгонки решения уравнения Ричардса к предписанным значениям влажности.

Работа поддержана грантом РФФИ № 15-07-08952.

- [1] *Засухин С. В., Засухина Е. С.* Идентификация параметров модели // Machine Learning and Data Analysis, Москва: Вычислительный центр им. А.А. Дородницына Российской академии наук, 2018.

Identification of model parameters

*Zasukhin Sergey*¹

s.zasukhin@yandex.ru

Zasukhina Elena^{2*}

elenazet2017@yandex.ru

¹Dolgoprudny, Moscow Institute of Physics and Technology

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

Exact knowledge of soil parameters is of critical importance in modeling water flow and transfer of dissolved substances in the aeration zone, for assessment of the danger of catastrophic floods. Their experimental determination is a difficult task. Alternative methods for finding parameters are the pedotransfer functions of Rosetta, obtained by the neural networks method, and inverse modeling approach. The first requires a large amount of input data. As a rule, the parameters determination by the second method performed on the basis of observed water retention curve and hydraulic conductivity. In the present paper, a model of water movement in unsaturated soil is considered. The movement of water is described by Richards equation. The parameters determination problem is formulated as an inverse modeling problem in which the desired parameters are determined as a result of fitting simulated values of soil moisture to their observed values. The numerical solution of the problem is carried out by the steepest descent method. The exact values of the objective function gradient are calculated by formulas of fast automatic differentiation. Numerical results show that the proposed method allows to determine the parameters with good accuracy. A feature of the proposed method is that the parameters are determined as a result of fitting the solution of the Richards equation to the observed values of soil moisture. Another feature is application of exact formulas of fast automatic differentiation for computing the objective function gradient in optimization process.

This research is funded by RFBR, grant 15-07-08952.

- [1] *Zasukhin S., Zasukhina E.* Identification of model parameters // Machine Learning and Data Analysis, Moscow: Dorodnicyn Computing Centre of the Russian Academy of Sciences, 2018.

Архитектура Mathbrain.ru - облачного решения для анализа данных энцефалографии

Рыкунов Станислав Дмитриевич^{1*} rykunov@impb.ru
*Оплачко Екатерина Сергеевна*¹ oplachkoe@gmail.com
*Устинин Михаил Николаевич*¹ u_m_n@mail.ru

¹Пуццино, ИМПБ РАН - филиал ИПМ им. М.В. Келдыша РАН

DOI: 10.30826/IDP201883

Mathbrain.ru - это частное облачное решение для хранения и анализа данных энцефалографии. Оно построено с использованием таких технологий, как Docker Swarm - система оркестровки облачных контейнеров, хранилище Minio S3, язык программирования Python для вычислений. Облачная реализация необходима из-за больших объемов данных и их высокой сложности, и исключает необходимость в установке дополнительного программного обеспечения на стороне пользователя. Выбор механизма оркестровки Docker Swarm позволяет при необходимости легко масштабировать систему, используя общедоступные облачные платформы, такие как Microsoft Azure, Google Compute Engine или Amazon EC2. Он предоставляет инструменты для быстрого развертывания вычислительных задач, управления ресурсами и обладает минимальным влиянием на производительность. Mathbrain.ru предоставляет следующие методы анализа данных: - прямое и обратное преобразование Фурье временных рядов; - анализ независимых компонент; - анализ главных компонент; - количественный анализ энцефалограмм; - решение обратной задачи МЭГ по спектральным компонентам, независимым компонентам или главным компонентам. Пользователь взаимодействует с Mathbrain через веб-интерфейс, который позволяет загружать данные, выбирать параметры для анализа и смотреть результаты с помощью графиков. Работа выполнена при поддержке Программы Президиума РАН №27 «Фундаментальные проблемы решения сложных практических задач с помощью суперкомпьютеров» и грантов РФФИ 17-29-02178, 18-29-02087, 16-07-00937, 16-07-01000, 17-07-00677, 17-07-00686.

- [1] *Oplachko E., Rykunov S., Ustinin M.* SaaS Platform for Time Series Data Handling. // EPJ Web of Conferences, 2018, vol. 173, — p. 05013 <https://doi.org/10.1051/epjconf/201817305013>

Architecture of Mathbrain.ru - cloud solution for encephalography data analysis

*Rykunov Stanislav*¹

rykunov@impb.ru

*Oplachko Ekaterina*¹

oplachkoe@gmail.com

*Ustinin Mikhail*¹*

u_m_n@mail.ru

¹Pushchino, IMPB RAS - Branch of KIAM RAS

Mathbrain.ru is a private cloud solution for storing and analyzing encephalography data. It is built using such technologies as Docker Swarm – a cloud container orchestration system, Minio S3 storage, Python programming language for computations. Cloud design is a necessity, due to large data volumes and high complexity. Also it eliminated the need for additional software installation on user's side. Choice of Docker swarm orchestration engine allows an easy scalability in case of need, using public cloud platforms, such as Microsoft Azure, Google Compute Engine or Amazon EC2. It also provides tools for rapid computational tasks scheduling and deployment, resource allocation and has a little to none performance overhead. Mathbrain.ru provides following data analysis methods: - forward and inverse Fourier analysis of multichannel time-series; - independent component analysis; - principal component analysis; - quantitative encephalogram analysis; - MEG inverse problem solution from spectral components, independent components or principal components. User interacts with Mathbrain through web interface which allows to upload and download data, select parameters for analysis and review results using interactive plots. This research is funded by Programs of the Presidium of the Russian Academy of Sciences 27 «Fundamental problems of solving complex practical problems with the help of supercomputers» and RFBR, grants 17-29-02178, 18-29-02087, 16-07-00937, 16-07-01000, 17-07-00677, 17-07-00686.

- [1] *Oplachko E., Rykunov S., Ustinin M.* SaaS Platform for Time Series Data Handling. // EPJ Web of Conferences, 2018, vol. 173, — p. 05013 <https://doi.org/10.1051/epjconf/201817305013>

О вероятности и времени инверсии большого спина

Карацуба Екатерина Анатольевна^{1*} ekaratsuba@gmail.com
*Моретти Паоло*² pmoretti56@gmail.com

¹Москва, Федеральный исследовательский центр "Информатика и управление" Российской Академии Наук

²Флоренция, Институт комплексных систем

DOI: 10.30826/IDP201884

Коэффициент $|b_n(t)|^2$, который определяется как вероятность найти систему в n -ом спиновом состоянии, был вычислен и оказался некоторой сложной тригонометрической суммой. Примерное выражение для $b_n(t)$ посредством соответствующих функций Бесселя было получено только при условии, что $n \ll N$, $at \ll N$ (N есть общее число спиновых состояний, α – параметр), поскольку только в этом случае можно обоснованно заменить тригонометрическую сумму по всем спиновым состояниям соответствующим интегралом. Используя эвристические рассуждения, это соотношение предполагается справедливым также при $n = N$, что соответствует переходу от состояния спин-вверх к состоянию спин-вниз, и таким образом можно вычислить время инверсии. Оно определяется первым максимумом функции $|b_N(t)|^2$. Однако из-за значимости момента времени инверсии, связывающего теорию с экспериментальными данными, нужны были более точные и обоснованные вычисления. Настоящий доклад посвящён исследованию $b_N(t)$, проведенному в [1], где для амплитуды вероятности $b_N(t)$ была выведена новая формула через функции Бесселя с большими индексами, при $N \geq 2$ и для любого t . Получено асимптотически точное выражение для времени инверсии спина. Применение этой полезной формулы показано на примере кластера Fe8.

Работа первого автора поддержана грантами РФФИ РФФИ 13-01-00657 А и РФФИ-РЖД:13-07-13136 офм_м_РЖД.

- [1] *Карацуба Е. А., Моретти П.* Вероятность инверсии большого спина в форме асимптотического разложения в ряды функций Бесселя. // Математическое моделирование, 2018 (на рецензии).

On the probability and time of the inversion of a large spin

*Ekatherina Karatsuba*¹★

author.ekaratsuba@gmail.com

*Paolo Moretti*²

pmoretti56@gmail.com

¹Research Center 'Informatics and Control' of RAS

²Via Madonna del Piano 10 - 50019 Sesto Fiorentino, Florence - Italy

The coefficient $|b_n(t)|^2$, that gives the probability of finding the system in the n -th spin state, has been calculated and it is found to be a somewhat involute trigonometric sum. An approximate expression for $b_n(t)$ by means of the corresponding Bessel functions was obtained only if $n \ll N$, $\alpha t \ll N$ (N is the total number of spin states, α is a parameter), since the trigonometric sum which runs through all the spin states can be replaced by an integral in a careful way only in this case. Using a heuristic argumentation, this form has been assumed to be valid also for $n = N$, which corresponds to the transition from spin-up to spin-down, and the inversion time has been calculated in this way. It is given by the first maximum of $|b_N(t)|^2$. However, due to the important role of the inversion time, which connects the theory with the experimental spectra, a better calculation is in order.

The present talk is about investigations performed in [1] where a new formula for the probability amplitude $b_N(t)$, for $N \geq 2$ and for any t , was derived in terms of the Bessel functions with large indexes. An asymptotically exact expression is obtained for the spin inversion time. An application of this useful formula is shown in the example of the Fe8 cluster.

- [1] *Karatsuba E., Moretti P.* The inversion probability of a large spin as an asymptotic expansion in series of Bessel functions // *Mathematical Models and Computer Simulations*, 2018 (submitted to a journal).

Эвристический подход к тестированию нейронных сетей на основе концепций тестирования программного обеспечения

Карпов Юрий Леонидович^{1*} y.l.karpov@yandex.ru

Карпов Леонид Евгеньевич^{2,3} mak@ispras.ru

Сметанин Юрий Геннадьевич^{4,5} yismetanin@rambler.ru

¹Москва, Люкссофт Россия

²Москва, Институт системного программирования имени

В. П. Иванникова РАН

³Москва, Московский Государственный университет имени

М. В. Ломоносова

⁴Москва, Федеральный исследовательский центр «Информатика и управление» РАН

⁵Москва, Московский Физико-Технический институт (технический университет)

DOI: 10.30826/IDP201885

На практике выбор архитектуры нейронной сети при решении конкретной задачи производится наугад, на основе опыта разработчиков. При построении глубоких нейронных сетей это может привести к выбору, сильно отличающемуся от наилучшего. Отличие нейронных сетей от программных систем, реализующих алгоритмические методы решения задач, достаточно существенно. При программной реализации алгоритмов можно проследить движение данных, тестирование программной системы может помочь обнаружить ошибки. В противоположность этому нейронная сеть иногда воспринимается в виде чёрного ящика и вывод о правильности работы делается на основании пар (вход, выход). Однако это не всегда так. Структура нейронной сети, используемой при решении конкретной задачи, может быть хорошо известна. Более того, хотя алгоритмическая связь элементов этой структуры друг с другом не всегда ясна, зачастую хорошо известны некоторые параметры структурных элементов, коррекция ошибок работы которых можно проводить примерно также, как проводятся принципиально эвристические процедуры тестирования и отладки программных систем.

Heuristic approach to testing neural networks that meets software testing concepts

Karpov Yury^{1,2*}

y.l.karpov@yandex.ru

*Karpov Leonid*²

mak@ispras.ru

*Smetanin Yury*²

ysmetanin@rambler.ru

¹Moscow, Luxoft Professional LLC

²Moscow, V. P. Ivannikov Institute for System Programming of Russian Academy of Sciences

³Moscow, Moscow State University

⁴Moscow, Federal Research Center “Informatics and Control” of Russian Academy of Sciences

⁵Moscow, Moscow Institute of Physics and Technology (Technical University)

In practice, the choice of the neural network architecture for solving a specific problem is based on the experience of developers. When constructing deep neural networks, this can lead to a choice that is far from the best.

The difference between neural networks and software systems that implement algorithmic methods for solving problems is quite significant. In software implementation of algorithms, one can trace the movement of data, testing the software system can detect and localize errors. In contrast, a neural network is often considered as a black box, and the conclusion about correctness of operations is based on its $\langle \text{input}, \text{output} \rangle$ pairs. However, this is not always the case. The structure of the neural network used to solve a particular problem can be well known a priori. Moreover, although the algorithmic connection among the elements of this structure is not always clear, some parameters of structural elements can be corrected in the same way as in essentially heuristic procedures for testing and debugging software systems.

In applied problems, type of input and output data and its dimension are determined by the specific nature of the problem and by the particular system within which the problem is being solved.

В конкретных применениях размерность и тип входных и выходных данных определяются спецификой задачи и конкретной системой, в рамках которой задача решается. Опираясь на эти данные создатели нейросетевых систем выбирают количество и типы нейронов, их разбиение на слои, виды этих слоев, способы взаимодействия между нейронами, веса синапсов, количество обучающих примеров, степень их влияния на выходные векторы, метрики близости входных и выходных векторов, алгоритмы построения матриц весов связей и другие характеристики. При этом наиболее продуктивным представляется подход, организованный на базе моделирования нейронной сети в виде обобщённого иерархического недетерминированного автомата, допускающего спонтанные изменения своего состояния. Этот подход, хорошо соотносящийся с принятыми нормами и стандартами, разработанными для поддержки процессов тестирования программного обеспечения, в настоящее время реализуется авторами в рамках проекта, поддержанного грантом РФФИ № 18-07-00697. Полная версия статьи была опубликована в журнале «Programming and Computer Software» [1].

- [1] *Karpov Yu., Karpov L., Smetanin Yu.* Heuristic approach to testing neural networks that meets software testin concepts // Programming and Computer Software, 2018. Vol. 44. No. 5. P. 5–25. <https://doi.org/10.1134/S0361768818050031>.

Based on these initial data, the designers of neural network systems select the number and types of neurons, their partition into layers, the types of these layers, the methods of interaction between neurons, the weight of the synapses, the activation functions, and some other characteristics.

The most promising seems the approach based on modeling a neural network by a generalized hierarchical nondeterministic automaton with spontaneous changes of its states. This approach complies with the accepted norms and standards for testing software systems. It is now implementing by authors in their research project supported by Russian Foundation for Basic Research (grant No. 18-07-00697).

Full version of the article was published in 'Programming and Computer Software' journal [1].

- [1] *Karpov Yu., Karpov L., Smetanin Yu.* Heuristic approach to testing neural networks that meets software testin concepts // Programming and Computer Software, 2018. Vol.44. No.5. P.5–25. <https://doi.org/10.1134/S0361768818050031>.

Разработка средств объектно-ориентированного логического программирования для слияния и совместного анализа терагерцового и трёхмерного видеоизображений

| | |
|--|------------------------|
| <i>Морозов Алексей Александрович</i> ^{1*} | morozov@cplire.ru |
| <i>Сушкова Ольга Сергеевна</i> ¹ | o.sushkova@mail.ru |
| <i>Полупанов Александр Фёдорович</i> ¹ | sashap55@mail.ru |
| <i>Анциперов Вячеслав Евгеньевич</i> ^{1,2} | antciperov@gmail.com |
| <i>Мансуров Геннадий Константинович</i> ¹ | gmansurov@mail.ru |
| <i>Папроцкий Станислав Константинович</i> ¹ | s.paprotskiy@gmail.com |
| <i>Янушко Александр Владимирович</i> ¹ | yanushko@gmail.com |
| <i>Петрова Надежда Геннадиевна</i> ¹ | petrova@cplire.ru |
| <i>Бугаёв Александр Степанович</i> ^{1,2} | bugaev@cplire.ru |

¹Москва, ИРЭ им. В. А. Котельникова РАН

²Долгопрудный, Московский физико-технический институт

DOI: 10.30826/IDP201886

Целью работы является создание методов и средств терагерцового интеллектуального видеонаблюдения, то есть, автоматического анализа видеоизображений в терагерцовом диапазоне частот. Терагерцовое видеонаблюдение открывает новые возможности в обеспечении безопасности, в частности, оно позволяет дистанционно выявлять опасные предметы, скрытые под одеждой. Вместе с тем, особенности терагерцового видеоизображения, такие как низкое разрешение, слабый контраст и низкое соотношение сигнал/шум, приводят к необходимости разработки новых методов и подходов к автоматическому анализу видеоизображений. В статье описаны экспериментальные средства объектно-ориентированного логического программирования, разработанные для решения задач трёхмерного видеоизображений. Работа поддержана РФФИ (проект №16-29-09626-офи-м).

- [1] Морозов А. А., Сушкова О. С., Полупанов А. Ф., Анциперов В. Е., Мансуров Г. К., Папроцкий С. К., Янушко А. В., Петрова Н. Г., Бугаёв А. С. Разработка средств объектно-ориентированного логического программирования для слияния и совместного анализа терагерцового и трёхмерного видеоизображений // РЭНСИТ. — 2018. — Т. 10, № 2. — С. 265–274.

Development of means of object-oriented logic programming for fusion and joint analysis of terahertz and 3D video images

Alexei Morozov^{1*}

morozov@cplire.ru

*Olga Sushkova*¹

o.sushkova@mail.ru

*Alexander Polupanov*¹

sashap55@mail.ru

Viacheslav Antsiperov^{1,2}

antciperov@gmail.com

*Gennady Mansurov*¹

gmansurov@mail.ru

*Stanislav Paprotskiy*¹

s.paprotskiy@gmail.com

*Alexander Yanushko*¹

yanushko@gmail.com

*Nadezda Petrova*¹

petrova@cplire.ru

Alexander Bugaev^{1,2}

bugaev@cplire.ru

¹Moscow, Kotel'nikov IRE RAS

²Dolgoprudny, MIPT

The aim of the work is to create methods and software tools for the terahertz range intelligent video surveillance, that is, for automatic analysis of video images in the terahertz frequency range. Terahertz video surveillance provides unique prospects in the area of public safety; in particular, it allows you to remotely identify weapons and dangerous items hidden under clothing on the human body. However, such characteristics of terahertz video, as low resolution, low contrast, and low signal-to-noise ratio, lead to the need to develop new approaches to automatic video analysis. A new method of the terahertz video surveillance based on the fusion of the terahertz video with 3D video is proposed. New means of the object-oriented logic programming for the semantic fusion of the terahertz and 3D video images are described. The developed method provides a real-time fusion of the terahertz video acquired using the THERZ-7A (Astrohn Technology Ltd) subterahertz scanning device (0.23-0.27 THz) and 3D video data acquired using the Kinect 2 time-of-flight (ToF) sensor.

The research was supported by the Russian Foundation for Basic Research (grant 16-29-09626-ofi-m).

- [1] *Morozov A. A., Sushkova O. S., Polupanov A. F., Antsiperov V. E., Mansurov G. K., Paprotskiy S. K., Yanushko A. V., Petrova N. G., Bugaev A. S.* Development of means of object-oriented logic programming for fusion and joint analysis of terahertz and 3D video images // RENSIT. — 2018. — V. 10, No 2. — p. 265–274.

Содержание

| | |
|---|-----------|
| Теория и методы машинного обучения | 10 |
| <i>Бериков В. Б., Новиков И. А.</i> | |
| Модель и вычислительно эффективный метод ансамблевой кластеризации | 10 |
| <i>Дюкова Е. В., Масляков Г. О., Прокофьев П. А.</i> | |
| Задача дуализации над произведением цепей: асимптотика типичного числа решений | 12 |
| <i>Двоенко С. Д., Пшеничный Д. О.</i> | |
| Применение t-распределения Стьюдента при метрической коррекции матриц парных сравнений | 16 |
| <i>Неделько В. М.</i> | |
| Статистические критерии согласия на основе оценки скользящего экзамена | 18 |
| <i>Владимирова М. Р., Арбель Дж.</i> | |
| Исследование свойств распределений байесовских нейронных сетей | 20 |
| <i>Зубюк А. В.</i> | |
| Новый подход к специфичности возможностей мер и его роль в теории принятия решений | 22 |
| <i>Пытьев Ю. П., Фаломкина О. В., Чуличков А. И., Шиликин С. А.</i> | |
| Математический формализм субъективного моделирования | 24 |
| <i>Нейчев Р. Г., Стрижов В. В.</i> | |
| Информативные априорные предположения в задаче привилегированного обучения | 26 |
| <i>Балтеев О. Ю., Стрижов В. В.</i> | |
| Байесовский выбор наиболее правдоподобной структуры модели глубокого обучения | 28 |
| <i>Носова С. А., Турлапов В. Е.</i> | |
| GLCM, kNN and Meanshift в задаче детектирования нейронов по изображениям срезов мозга, окрашенных по Нисслю | 30 |

| | |
|---|-----------|
| <i>Уткин Л. В., Мелдо А. А.</i> Весовой случайный лес для построения управляемых моделей выживаемости | 32 |
| <i>Мелдо А. А., Уткин Л. В.</i> Автоматизированная система дифференциальной диагностики заболеваний лёгких | 34 |
| <i>Грбовой А. В., Базтеев О. Ю., Стрижов В. В.</i> Определение релевантности параметров нейросети методом Белсли | 36 |
| <i>Макарова А. И., Сулимова В. В.</i> Быстрое приближенное обучение двухклассовому распознаванию по методу опорных векторов в условиях большого числа объектов | 38 |
| <i>Яковлев П. А.</i> Метод быстрого множественного попарного выравнивания на основе префиксных деревьев | 40 |
| Дискретная оптимизация и сложность вычислений | 42 |
| <i>Андрианов А. Н., Горнов А. Ю., Аникин А. С.</i> Вычислительные технологии решения квазисепарабельных задач безусловной минимизации | 42 |
| <i>Гимади Э. Х., Курочкина А. А., Нагорная Е. А., Шамардин Ю. В., Шевяков А. С.</i> О некоторых полиномиально разрешимых подклассах сетевой задачи размещения | 44 |
| <i>Гимади Э. Х., Курочкина А. А., Цидулко О. Ю.</i> О сетевой задаче размещения с ограничениями на пропускные способности и объемы производства | 46 |
| <i>Гимади Э. Х., Истомин А. М., Шин Е. Ю.</i> О задаче Random MST с верхним ограничением на диаметр остова | 48 |
| <i>Гимади Э. Х., Гончаров Е. Н., Мишин Д. В.</i> О реализации точного метода решения задачи объемно-календарного планирования сетевого проекта с ограниченными ресурсами | 50 |

| | |
|--|----|
| <i>Горнов А. Ю.</i> | |
| Алгоритмы облачной аппроксимации невыпуклых множеств в конечномерном пространстве | 52 |
| <i>Кельманов А. В., Хандеев В. И., Панасенко А. В.</i> | |
| Точные алгоритмы для специальных случаев двух труднорешаемых задач 2-кластеризации | 54 |
| <i>Кельманов А. В., Хандеев В. И., Панасенко А. В.</i> | |
| Рандомизированные алгоритмы для некоторых задач кластеризации | 56 |
| <i>Кельманов А. В., Хамидуллин С. А., Хандеев В. И., Пяткин А. В.</i> | |
| Точные алгоритмы для двух задач поиска наибольшего подмножества и наибольшей подпоследовательности | 58 |
| <i>Кельманов А. В., Пяткин А. В., Хамидуллин С. А., Хандеев В. И.</i> | |
| Точный алгоритм поиска кластера наибольшего размера в задаче 2-кластеризации целочисленной последовательности | 60 |
| <i>Кельманов А. В., Пяткин А. В., Хамидуллин С. А., Хандеев В. И., Шенмайер В. В., Шамардин Ю. В.</i> | |
| Приближенный полиномиальный алгоритм для задачи поиска подпоследовательности наибольшей длины в конечной последовательности точек евклидова пространства | 62 |
| <i>Кельманов А. В., Пяткин А. В., Хандеев В. И.</i> | |
| NP-трудность некоторых максиминных задач кластеризации | 64 |
| <i>Кельманов А. В., Михайлова Л. В., Романченко С. М.</i> | |
| Об одной задаче суммирования элементов, выбираемых из семейства конечных числовых последовательностей | 66 |
| <i>Хачай М. Ю., Огородников Ю. Ю.</i> | |
| Улучшенная полиномиальная приближённая схема для задачи маршрутизации транспорта с ограничениями на грузоподъёмность и временные промежутки обслуживания | 68 |
| <i>Хачай М. Ю., Хачай Д. М.</i> | |
| Неулучшаемая гарантированная оценка точности для задачи о k медианах на отрезке | 70 |

| | |
|--|-----------|
| <i>Хачай М. Ю., Незнахина Е. Д.</i> Разрешимость евклидовой обобщенной задачи коммивояжера на сетке фиксированной высоты | 72 |
| <i>Зароднюк Т. С., Горнов А. Ю.</i> Поиск глобального минимума функционала на основе технологий обучения с подкреплением | 74 |
| <i>Рогозин А. В., Урибе С., Гасников А. В., Мальковский Н. В., Недичь А.</i> Ускоренный метод для децентрализованной распределённой оптимизации на меняющихся со временем сетях . . . | 76 |
| <i>Иванова А. С., Двуреченский П. Е., Гасников А. В.</i> Задача распределения ресурсов с точки зрения композитной оптимизации | 78 |
| <i>Бекларян Л. А., Бекларян А. Л.</i> К вопросу существования солитонных решений для систем с полиномиальным потенциалом и их численная реализация | 80 |
| Анализ сигналов и временных рядов | 82 |
| <i>Анциперов В. Е.</i> Идентификация сигналов на основе максимально правдоподобного соответствия их случайной выборки стохастическим описаниям прецедентов | 82 |
| <i>Филипенков Н. В., Петрова М. А.</i> Поиск плавно меняющихся пространственных закономерностей | 84 |
| <i>Исаченко Р. В., Стрижов В. В.</i> Снижение размерности с помощью проекции на скрытое пространство в задаче декодирования сигналов | 86 |
| <i>Мандрикова О. В., Фетисова Н. В., Полозов Ю. А.</i> Моделирование и анализ природных временных рядов на основе обобщенной многокомпонентной модели | 88 |

| | |
|--|------------|
| <i>Мандрикова О. В., Залаяев Т. Л., Геппенер В. В., Мандрикова Б. С.</i> Анализ данных нейтронных мониторов и выделение спорадических эффектов на основе нейронных сетей и вейвлет-преобразования | 90 |
| <i>Манило Л. А.</i> Метод анализа псевдофазового портрета при распознавании биосигналов с нелинейными свойствами | 92 |
| <i>Панкратов А. Н., Тетуев Р. К., Пятков М. И.</i> Оптимизация параллельного алгоритма глобального выравнивания протяженных последовательностей | 94 |
| <i>Панкратова Н. М., Рыкунов С. Д., Бойко А. И., Молчанова Д. А., Устинин М. Н.</i> Спектральные и пространственные особенности энцефалограмм при психических расстройствах | 96 |
| <i>Кириллюк И. Л., Сенько О. В.</i> Верификация регрессионных моделей во временных рядах | 98 |
| <i>Моттль В. В., Красоткина О. В., Морозов А. О.</i> Оценивание состава инвестиционного портфеля в большом множестве биржевых активов | 100 |
| <i>Моттль В. В., Красоткина О. В., Черноусова Е. О., Рыбка Е. М.</i> Формирование инвестиционного портфеля по неточной информации о доходностях активов | 102 |
| <i>Моттль В. В., Красоткина О. В., Морозов А. О., Медведев Е. М.</i> Алгоритмическая реализация методологии оценивания состава инвестиционных портфелей | 104 |
| Обработка изображений | 106 |
| <i>Бакланов А. П., Пасынков М. К., Хачай М. Ю.</i> Семантическая сегментация аэрокосмических изображений промышленных пальмовых плантаций с использованием вполне конволюционных сверточных нейронных сетей | 106 |

| | |
|--|------------|
| <i>Гнеушев А. Н., Самсонов Н. А.</i> Метод проекций в пространстве лучевого преобразования Радона для детектирования пешеходов | 108 |
| <i>Грачева И. А., Копылов А. В.</i> Алгоритм сжатия динамического диапазона HDR изображений на основе фильтрации с сохранением структуры | 110 |
| <i>Ханыков И. Г.</i> Модель скоростной кластеризации пикселей изображения на основе метода Уорда | 112 |
| <i>Федотова С. А., Середин О. С., Кушмир О. А.</i> Об использовании инвариантов отражения в практическом решении задачи определения зеркальной симметрии | 114 |
| <i>Лемтюжникова Д. В.</i> Задача распознавания символов в художественных изображениях | 116 |
| <i>Мурашов Д. М., Березин А. В., Иванова Е. Ю.</i> Измерение плотности холста картин по изображениям . . | 118 |
| <i>Мурынин А. Б., Бондур В. Г.</i> Измерение пространственных спектров морских волн по оптическим аэрокосмическим изображениям высокого разрешения | 122 |
| <i>Мурынин А. Б., Гороховский К. Ю., Игнатьев В. Ю., Ракова К. О.</i> Поиск оптимальных параметров вероятностного алгоритма повышения пространственного разрешения мультиспектральных спутниковых изображений | 124 |
| Анализ и распознавание изображений | 126 |
| <i>Андрянов Н. А., Васильев К. К.</i> Разработка и исследование дважды стохастических моделей случайных полей для представления и обработки изображений со сложной структурой | 126 |
| <i>Ефимов Ю. С., Одиноких Г. А., Соломатин И. А., Коробкин М. В., Матвеев И. А.</i> Метод детектирования подделок в мобильной системе распознавания по радужной оболочке | 128 |

| | |
|--|------------|
| <i>Лебедев М. А., Коमारов Д. В., Выголов О. В., Визильтер Ю. В.</i> Выделение отличий на разносезонных изображениях земной поверхности при помощи генеративных состязательных сетей | 130 |
| <i>Новик В. П., Матвеев И. А.</i> Сравнение информативных признаков изображений радужки глаза методом оптимального пути | 132 |
| <i>Федотов Н. Г., Сёмов А. А., Моисеев А. В.</i> Теоретические основы гипертрейс-преобразования | 134 |
| <i>Васильев К. К., Дементьев В. Е., Андриянов Н. А.</i> Анализ эффективности нелинейной фильтрации многозональных спутниковых изображений и их последовательностей | 136 |
| Морфология изображений | 138 |
| <i>Одиноких Г. А., Коробкин М. В., Ефимов Ю. С., Соломатин И. А., Матвеев И. А.</i> Метод выделения радужки на изображении, полученном в сложных условиях окружения | 138 |
| <i>Середин О. С., Копылов А. В., Родионов Д. С.</i> Использование Microsoft Kinect V2 для детектирования падений человека | 140 |
| <i>Соломатин И. А., Одиноких Г. А., Ефимов Ю. С., Матвеев И. А.</i> Аппроксимация границ радужки классифицирующей свёрточной нейронной сетью | 142 |
| Анализ текстов и информационный поиск | 144 |
| <i>Емельянов Г. М., Михайлов Д. В., Козлов А. П.</i> Мера TF-IDF, сила связи слов и эталон языкового описания единицы знаний в открытых тестах | 144 |
| <i>Янина А. О., Воронцов К. В.</i> Тематический информационный поиск | 146 |

| | |
|--|-----|
| <i>Жариков И. Н., Апишев М. А., Воронцов К. В.</i> Гиперграфовые многомодальные вероятностные тематические модели транзакционных данных | 148 |
| <i>Алешин И. С.</i> Формализация задач поиска сгущений в разреженных булевых матрицах | 150 |
| <i>Ланге М. М., Ланге А. М.</i> Модель классификации на основе средней взаимной информации | 152 |
| Анализ биомедицинских сигналов | 156 |
| <i>Устинин М. Н., Рыкунов С. Д., Рыкунова Е. Д., Бойко А. И.</i> Методы изучения локальных свойств функциональной структуры человеческого тела | 156 |
| <i>Сушкова О. С., Морозов А. А., Габова А. В., Карабанов А. В.</i> Исследование проблемы множественного сравнения при анализе всплескообразной электрической активности мозга и мышц | 158 |
| <i>Янковская А. Е., Пеккер Я. С., Бразовский К. С.</i> Основанная на тестовых методах распознавания образов прикладная интеллектуальная система определения состояния пробуждения пациента во время общей анестезии | 160 |
| <i>Янковская А. Е., Гергет О. М.</i> Конвергенция нескольких методов для интеллектуального анализа данных и знаний | 162 |
| <i>Сулимова В. В., Красоткина О. В., Бухонов С. А., Моттль В. В., Уиндридж Д.</i> Обнаружение отклонений на маммограммах с использованием интерфейса мозг-компьютер | 164 |
| Биоинформатика | 166 |
| <i>Тихонов Д. А., Куликова Л. И., Ефимов А. В.</i> Анализ торсионных углов между осями спиралей в спиральных парах белковых молекул | 166 |

| | |
|--|------------|
| <i>Васильев Е. П., Комура Т., Турлапов В. Е., Никольский А. В.</i> Жестовое управление дополненной реальностью при кардиологических интервенциях | 168 |
| Геоинформатика | 170 |
| <i>Флоринский И. В., Филиппов С. В., Абрамова А. С., Зарай- ская Ю. А., Селезнева Е. В.</i> Разработка системы для моделирования рельефа дна Се- верного Ледовитого океана | 170 |
| <i>Визильтер Ю. В., Горбачевич В. С., Хайдаров А. С., Яко- влев В. А.</i> Синтез кривых ГИС с использованием генеративных конкурирующих глубоких конволюционных нейронных сетей | 172 |
| <i>Хайретдинов М. С., Агафонов В. М., Ковалевский В. В., Воскобойникова Г. М.</i> Многофакторный метод геоэкологоохранного прогнози- рования | 174 |
| <i>Наумов В. А., Нелюбина Е. А., Рязанов В. В., Виногра- дов А. П.</i> Анализ и прогнозирование гидрологического ряда на ос- нове обобщённых прецедентов | 178 |
| <i>Засухин С. В., Засухина Е. С.</i> Идентификация параметров модели | 180 |
| Прикладные системы | 182 |
| <i>Рыжунев С. Д., Оплачко Е. С., Устинин М. Н.</i> Архитектура Mathbrain.ru - облачного решения для ана- лиза данных энцефалографии | 182 |
| <i>Карацуба Е. А., Моретти П.</i> О вероятности и времени инверсии большого спина | 184 |
| Программная инженерия | 186 |

| | |
|---|-----|
| <i>Карпов Ю. Л., Карпов Л. Е., Сметанин Ю. Г.</i> Эвристический подход к тестированию нейронных сетей на основе концепций тестирования программного обеспе- чения | 186 |
| <i>Морозов А. А., Сушкова О. С., Полупанов А. Ф., Анци- перов В. Е., Мансуров Г. К., Папроцкий С. К., Януш- ко А. В., Петрова Н. Г., Бугаёв А. С.</i> Разработка средств объектно-ориентированного логиче- ского программирования для слияния и совместного ана- лиза терагерцового и трёхмерного видеоизображений . . | 190 |
| Авторский указатель | 194 |

Contents

| | |
|--|----|
| Machine Learning | 10 |
| <i>Berikov V., Novikov I.</i> | |
| Model and computationally efficient method of ensemble clustering | 11 |
| <i>Djukova E., Masliakov G., Prokofyev P.</i> | |
| The problem of dualization over the product of chains: asymptotics for a typical number of solutions | 13 |
| <i>Dvoenko S., Pshenichny D.</i> | |
| Using of Student's t-distribution for metric correction of pairwise comparison matrices | 17 |
| <i>Nedelko V</i> | |
| Statistical Tests Based on Cross-Validation Estimation | 19 |
| <i>Vladimirova M., Arbel J.</i> | |
| Distributional properties of Bayesian neural networks | 21 |
| <i>Zubyuk A.</i> | |
| A new approach to specificity in possibility theory: Decision-making point of view | 23 |
| <i>Pytyev Y., Falomkina O., Chelichkov A., Shiskin S.</i> | |
| Mathematical formalism of subjective modeling | 25 |
| <i>Neychev R., Strijov V.</i> | |
| Informative prior in privileged learning problems | 27 |
| <i>Bakhteev O., Strijov V.</i> | |
| Bayesian deep learning optimal model structure selection | 29 |
| <i>Nosova S., Turlapov V.</i> | |
| GLCM, kNN and Meanshift for neuron detection on Nissl-stained brain slice image | 31 |
| <i>Utkin L.V., Meldo A.A.</i> | |
| A weighted random survival forest for constructing controllable models | 33 |
| <i>Meldo A.A., Utkin L.V.</i> | |
| A computer-aided system for differential diagnosis of lung diseases | 35 |

| | |
|--|-----------|
| <i>Grabovoy A., Bakhteev O., Strijov V.</i> Assessment of the relevance of the Neural network parameters with Belsly method | 37 |
| <i>Makarova A., Sulimova V.</i> Fast approximate two-class SVM learning in the case of a large number of objects | 39 |
| <i>Yakovlev P.</i> Fast Trie-based method for multiple pairwise sequence alignment | 41 |
| Discrete Optimization and Computational Complexity | 42 |
| <i>Andrianov A., Gornov A., Anikin A.</i> Computational technologies for solving quasi-separable problems of unconditional minimization | 43 |
| <i>Gimadi E., Kurochkina A., Nagornaya E., Shamardin Y., Shevyakov A.</i> On some polynomially solvable subclasses of Network FLP . | 45 |
| <i>Gimadi E., Kurochkina A., Tsidulko O.</i> On the Network Restricted Capacitated Facility Location Problem | 47 |
| <i>Gimadi E., Istomin A., Shin E.</i> On Bounded Diameter MST Problem on Random Instances | 49 |
| <i>Gimadi E., Goncharov E., Mishin D.</i> On Some Realizations of Solving the Resource Constrained Project Scheduling Problems | 51 |
| <i>Gornov A.</i> Algorithms of cloud approximation for nonconvex sets in a finite-dimensional space | 53 |
| <i>Kelmanov A., Khandeev V., Panasenko A.</i> Exact Algorithms for the Special Cases of Two Hard to Solve Problems | 55 |
| <i>Kelmanov A., Khandeev V., Panasenko A.</i> Randomized Algorithms for Some Clustering Problems . . . | 57 |
| <i>Kelmanov A., Khamidullin S., Khandeev V., Pyatkin A.</i> Exact Algorithms for Two Problems of Searching for the Largest Subset and Longest Subsequence | 59 |

| | |
|---|-----------|
| <i>Kelmanov A., Pyatkin A., Khamidullin S., Khandeev V.</i> An Exact Algorithm of Searching for the Largest Size Cluster in an Integer Sequence 2-Clustering Problem | 61 |
| <i>Kelmanov A., Pyatkin A., Khamidullin S., Khandeev V., Shenmaier V., Shamardin Y.</i> An Approximation Polynomial Algorithm for a Problem of Searching for the Longest Subsequence in a Finite Sequence of Points in Euclidean Space | 63 |
| <i>Kelmanov A., Pyatkin A., Khandeev V.</i> NP-hardness of Some Max-Min Clusterization Problems | 65 |
| <i>Kelmanov A., Mikhailova L., Romanchenko S.</i> On a Problem of Summing Elements Chosen from a Family of Finite Numerical Sequences | 67 |
| <i>Khachay M., Ogorodnikov I.</i> Improved Polynomial Time Approximation Scheme for Capacitated Vehicle Routing Problem with Time Windows | 69 |
| <i>Khachay M., Khachay D.</i> Attainable accuracy guarantee for the k -medians clustering in a line segment | 71 |
| <i>Khachay M., Neznakhina K.</i> Towards Tractability of the Euclidean Generalized Traveling Salesman Problem in Grid Clusters Defined by a Grid of Bounded Height | 73 |
| <i>Zarodnyuk T., Gornov A.</i> A global search of minimum functional based on reinforcement learning | 75 |
| <i>Rogozin A., Uribe C., Gasnikov A., Malkovsky N., Nédich A.</i> An accelerated method for decentralized distributed optimization on time-varying graphs | 77 |
| <i>Ivanova A., Dvurechensky P., Gasnikov A.</i> Composite optimization for the resource allocation problem | 79 |
| <i>Beklaryan L., Beklaryan A.</i> On the existence of soliton solutions for systems with a polynomial potential and their numerical realization | 81 |
| Signal and Time Series Analysis | 82 |

| | |
|---|-----|
| <i>Antsiperov V.</i> | |
| Signals identification based on the maximum-likelihood matching of their random samples to available precedents stochastic descriptions | 83 |
| <i>Filipenkov N., Petrova M.</i> | |
| Mining the Slightly Changing Patterns in Geospatial Data . | 85 |
| <i>Isachenko R., Strijov V.</i> | |
| Dimensionality Reduction for Multicorrelated Signal Decod- ing with Projections to Latent Space | 87 |
| <i>Mandrikova O., Fetisova N., Polozov Yu.</i> | |
| Modeling and analysis of natural time series on the basis of general multicomponent model | 89 |
| <i>Mandrikova O., Zalyaev T., Geppener V., Mandrikova B.</i> | |
| Modeling and analysis of natural time series on the basis of general multicomponent model | 91 |
| <i>Manilo L.</i> | |
| The method of analysis of pseudo-phase portrait in recogni- tion of biosignals with nonlinear properties | 93 |
| <i>Pankratov A., Tetuev R., Pyatkov M.</i> | |
| Optimization of parallel algorithm for global alignment of long sequences | 95 |
| <i>Pankratova N., Rykunov S., Boyko A., Molchanova D., Us- tinin M.</i> | |
| Spectral and spatial features of encephalograms in mental disorders | 97 |
| <i>Kirilyuk I., Senko O.</i> | |
| Verification of regression models in time series | 99 |
| <i>Mottl V., Krasotkina O., Morozov A.</i> | |
| Factor Search in Returns Based Analysis of Investment Port- folios | 101 |
| <i>Mottl V., Krasotkina O., Chernousova E., Rybka E.</i> | |
| Investment portfolio composition from uncertain forecast on stock market asset returns | 103 |
| <i>Mottl V., Krasotkina O., Morozov A., Medvedev A.</i> | |
| Algorithmic Implementation of Factor Search in Returns Based Analysis of Investment Portfolios | 105 |

| | |
|--|-----|
| Image Processing | 106 |
| <i>Baklanov A., Pasyukov M., Khachay M.</i> | |
| Semantic segmentation of industrial oil palm plantations using Fully Convolutional Neural Networks | 107 |
| <i>Gneushev A., Samsonov N.</i> | |
| The method of projections in the Radon ray transform space for pedestrian detection | 109 |
| <i>Gracheva I., Kopylov A.</i> | |
| Dynamic range compression algorithm of HDR images on the basis of the structure-preserving filtering | 111 |
| <i>Khanykov I.</i> | |
| The Model of the High-Speed Clustering of the Image Pixels on Ward's Method Basis | 113 |
| <i>Fedotova S., Seredin O., Kushnir O.</i> | |
| On practical solution to the problem of reflection symmetry detection using reflection invariants | 115 |
| <i>Lemtyuzhnikova D.</i> | |
| Towards recognition of symbols in painting images | 117 |
| <i>Murashov D., Berezin A., Ivanova E.</i> | |
| Measuring canvas density from images | 119 |
| <i>Murynin A., Bondur V.</i> | |
| Measurement of sea wave spatial spectra from high-resolution optical aerospace imagery | 123 |
| <i>Murynin A., Gorokhovskiy K., Ignatiev V., Rakova K.</i> | |
| Parameters Optimization of the Novel Probabilistic Algorithm for Improving Spatial Resolution of Multispectral Satellite Images | 125 |
| Image Analysis and Recognition | 126 |
| <i>Andriyanov N., Vasiliev K.</i> | |
| Development and investigation of doubly stochastic random fields for representation and processing images with complex structure | 127 |
| <i>Efimov Y., Odinokikh G., Solomatin I., Korobkin M., Matveev I.</i> | |
| Iris Anti-spoofing Solution for Mobile Biometric Applications | 129 |

| | |
|---|-----|
| <i>Lebedev M., Komarov D., Vygolov O., Vizilter Y.</i> Change detection in remote sensing images using conditional adversarial networks | 131 |
| <i>Novik V., Matveev I.</i> Matching of informative features of iris images with the optimal path method | 133 |
| <i>Fedotov N., Syemov A., Moiseev A.</i> Theoretical Foundations of Hypertrace-transform | 135 |
| <i>Vasiliev K., Dementiev V., Andriyanov N.</i> Analysis of nonlinear filtering efficiency in processing multi-zone satellite images and their sequences | 137 |
| Morphological Image Processing | 138 |
| <i>Odinokikh G., Korobkin M., Efimov Y., Solomatin I., Matveev I.</i> Iris Segmentation in Challenging Conditions | 139 |
| <i>Seredin O., Kopylov A., Rodionov D.</i> Microsoft Kinect v2 for fall detection | 141 |
| <i>Solomatin I., Odinokikh G., Efimov Y., Matveev I.</i> Iris boundaries approximation by classifying convolutional neural network | 143 |
| Text Analysis and Information Retrieval | 144 |
| <i>Emelyanov G., Mikhaylov D., Kozlov A.</i> TF-IDF, coupling strength of words and standard for lingual description of knowledge unit in open tests | 145 |
| <i>Ianina A., Vorontsov K.</i> Topic-based information retrieval | 147 |
| <i>Zharikov I., Apishev M., Vorontsov K.</i> Hypergraph multimodal probabilistic topic modeling of transaction data | 149 |
| <i>Aleshin I.</i> Formalization of tasks searching dense submatrices in boolean sparse matrices | 151 |
| <i>Lange M., Lange A.</i> A data classification model based on the average mutual information | 153 |

| | |
|---|-----|
| Biomedical Signal Analysis | 156 |
| <i>Ustinin M., Rykunov S., Rykunova E., Boyko A.</i> | |
| Methods to study local properties of the human body functional structure | 157 |
| <i>Sushkova O., Morozov A., Gabova A., Karabanov A.</i> | |
| Investigation of the multiple comparisons problem in the wave train electrical activity analysis of the brain and limbs | 159 |
| <i>Yankovskaya A., Pekker J., Brazovskii K.</i> | |
| Intelligent system based on test methods of pattern recognition to detect awareness state during general anesthesia . . . | 161 |
| <i>Yankovskaya A., Gerget O.</i> | |
| Convergence of several methods for intelligent analysis of data and knowledge | 163 |
| <i>Sulimova V., Krasotkina O., Buchonov S., Mottl V., Windridge D.</i> | |
| Detection of abnormalities in mammograms using the brain-computer interface | 165 |
| Bioinformatics | 166 |
| <i>Tikhonov D., Kulikova L., Efimov A.</i> | |
| Analysis of Torsion Angles between Helical Axes in Pairs of Helices in Protein Molecules | 167 |
| <i>Vasiliev E., Komuro T., Turlapov V., Nikolsky A.</i> | |
| One hand aerial gesture control for AR-based cardiac interventions | 169 |
| Geoinformatics | 170 |
| <i>Florinsky I., Filippov S., Abramova A., Zarayskaya Y., Selezneva E.</i> | |
| Development of a system for modelling the Arctic Ocean floor topography | 171 |
| <i>Vizilter Y., Gorbatshevich V., Khaydarov A., Yakovlev A.</i> | |
| Well LOG synthesis using generative adversarial CNN . . . | 173 |
| <i>Khairatdinov M., Agaphonov V., Kovalevsky V., Voskoboinikova G.</i> | |
| Multifactorial method of geocological prediction | 175 |

| | |
|---|-----|
| <i>Naumov V., Nelyubina E., Ryazanov V., Vinogradov A.</i> Analysis and prediction of hydrological series based on generalized precedents | 179 |
| <i>Zasukhin S., Zasukhina E.</i> Identification of model parameters | 181 |
| Applied Systems | 182 |
| <i>Rykunov S., Oplachko E., Ustinin M.</i> Architecture of Mathbrain.ru - cloud solution for encephalography data analysis | 183 |
| <i>Karatsuba E., Moretti P.</i> On the probability and time of the inversion of a large spin | 185 |
| Software Engineering | 186 |
| <i>Karpov Yu., Karpov L., Smetanin Yu.</i> Heuristic approach to testing neural networks that meets software testing concepts | 187 |
| <i>Morozov A., Sushkova O., Polupanov A., Antsiperov V., Mansurov G., Paprotskiy S., Yanushko A., Petrova N., Bugaev A.</i> Development of means of object-oriented logic programming for fusion and joint analysis of terahertz and 3D video images | 191 |
| Author index | 198 |

Авторский указатель

- А**
- Абрамова А. С., 170
 Агафонов В. М., 174
 Алешин И. С., 150
 Андрианов А. Н., 42
 Андриянов Н. А., .. 126, 136
 Аникин А. С., 42
 Анциперов В. Е., 82, 190
 Апишев М. А., 148
 Арбель Дж., 20
- Б**
- Бакланов А. П., 106
 Бахтеев О. Ю., 28, 36
 Бекларян А. Л., 80
 Бекларян Л. А., 80
 Березин А. В., 118
 Бериков В. Б., 10
 Бойко А. И., 96, 156
 Бондур В. Г., 122
 Бразовский К. С., 160
 Бугаёв А. С., 190
 Бухонов С. А., 164
- В**
- Васильев Е. П., 168
 Васильев К. К., 126, 136
 Визильтер Ю. В., .. 130, 172
 Виноградов А. П., 178
 Владимирова М. Р., 20
 Воронцов К. В., 146, 148
 Воскобойникова Г. М., .. 174
 Выголов О. В., 130
- Г**
- Габова А. В., 158
- Гасников А. В., 76, 78
 Гешпенер В. В., 90
 Гергет О. М., 162
 Гимади Э. Х., . 44, 46, 48, 50
 Гнеушев А. Н., 108
 Гончаров Е. Н., 50
 Горбацевич В. С., 172
 Горнов А. Ю., 42, 52, 74
 Гороховский К. Ю., 124
 Грабовой А. В., 36
 Грачева И. А., 110
- Д**
- Двоенко С. Д., 16
 Двуреченский П. Е., 78
 Дементьев В. Е., 136
 Дюкова Е. В., 12
- Е**
- Емельянов Г. М., 144
 Ефимов А. В., 166
 Ефимов Ю. С., 128, 138, 142
- Ж**
- Жариков И. Н., 148
- З**
- Заляев Т. Л., 90
 Зарайская Ю. А., 170
 Зароднюк Т. С., 74
 Засухин С. В., 180
 Засухина Е. С., 180
 Зубюк А. В., 22
- И**
- Иванова А. С., 78

Иванова Е. Ю., 118
Игнатьев В. Ю., 124
Исаченко Р. В., 86
Истомин А. М., 48

К

Карабанов А. В., 158
Карацуба Е. А., 184
Карпов Л. Е., 186
Карпов Ю. Л., 186
Кельманов А. В., . 54, 56, 58,
60, 62, 64, 66
Кириллюк И. Л., 98
Ковалевский В. В., 174
Козлов А. П., 144
Комаров Д. В., 130
Комура Т., 168
Копылов А. В., 110, 140
Коробкин М. В., . . . 128, 138
Красоткина О. В., . 100, 102,
104, 164
Куликова Л. И., 166
Курочкина А. А., 44, 46
Кушнир О. А., 114

Л

Ланге А. М., 152
Ланге М. М., 152
Лебедев М. А., 130
Лемтюжникова Д. В., . . . 116

М

Макарова А. И., 38
Мальковский Н. В., 76
Мандрикова Б. С., 90
Мандрикова О. В., . . . 88, 90
Манило Л. А., 92
Мансуров Г. К., 190

Масляков Г. О., 12
Матвеев И. А., 128, 132, 138,
142
Медведев Е. М., 104
Мелдо А. А., 32, 34
Михайлов Д. В., 144
Михайлова Л. В., 66
Мишин Д. В., 50
Моисеев А. В., 134
Молчанова Д. А., 96
Моретти П., 184
Морозов А. А., 158, 190
Морозов А. О., 100, 104
Мотгль В. В., . 100, 102, 104,
164
Мурашов Д. М., 118
Мурынин А. Б., . . . 122, 124

Н

Нагорная Е. А., 44
Наумов В. А., 178
Неделько В. М., 18
Недичь А., 76
Незнахина Е. Д., 72
Нейчев Р. Г., 26
Нелюбина Е. А., 178
Никольский А. В., 168
Новик В. П., 132
Новиков И. А., 10
Носова С. А., 30

О

Огородников Ю. Ю., 68
Одиноких Г. А., 128, 138, 142
Опlachко Е. С., 182

П

Панасенко А. В., 54, 56

Панкратов А. Н., 94
 Панкратова Н. М., 96
 Папроцкий С. К., 190
 Пасынков М. К., 106
 Пеккер Я. С., 160
 Петрова М. А., 84
 Петрова Н. Г., 190
 Полозов Ю. А., 88
 Полупанов А. Ф., 190
 Прокофьев П. А., 12
 Пшеничный Д. О., 16
 Пытьев Ю. П., 24
 Пяткин А. В., . 58, 60, 62, 64
 Пятков М. И., 94

Р

Ракова К. О., 124
 Рогозин А. В., 76
 Родионов Д. С., 140
 Романченко С. М., 66
 Рыбка Е. М., 102
 Рыкунов С. Д., . 96, 156, 182
 Рыкунова Е. Д., 156
 Рязанов В. В., 178

С

Самсонов Н. А., 108
 Селезнева Е. В., 170
 Сёмов А. А., 134
 Сенько О. В., 98
 Середин О. С., 114, 140
 Сметанин Ю. Г., 186
 Соломатин И. А., .. 128, 138,
 142
 Стрижов В. В., 26, 28, 36, 86
 Сулимова В. В., 38, 164
 Сушкова О. С., 158, 190

Т

Тетуев Р. К., 94
 Тихонов Д. А., 166
 Турлапов В. Е., 30, 168

У

Уиндридж Д., 164
 Урибе С., 76
 Устинин М. Н., . 96, 156, 182
 Уткин Л. В., 32, 34

Ф

Фаломкина О. В., 24
 Федотов Н. Г., 134
 Федотова С. А., 114
 Фетисова Н. В., 88
 Филипенков Н. В., 84
 Филиппов С. В., 170
 Флоринский И. В., 170

Х

Хайдаров А. С., 172
 Хайретдинов М. С., 174
 Хамидуллин С. А., 58, 60, 62
 Хандеев В. И., 54, 56, 58, 60,
 62, 64
 Ханьков И. Г., 112
 Хачай Д. М., 70
 Хачай М. Ю., 68, 70, 72, 106

Ц

Цидулко О. Ю., 46

Ч

Черноусова Е. О., 102
 Чуличков А. И., 24

Ш

Шамардин Ю. В., 44, 62

| | |
|-----------------------|----|
| Шевяков А. С., | 44 |
| Шенмайер В. В., | 62 |
| Шин Е. Ю., | 48 |
| Шишкин С. А., | 24 |

Я

| | |
|----------------------|----------|
| Яковлев В. А., | 172 |
| Яковлев П. А., | 40 |
| Янина А. О., | 146 |
| Янковская А. Е., ... | 160, 162 |
| Янушко А. В., | 190 |

Author index

- A**
- Abramova A., 171
 Agaphonov V., 175
 Aleshin I., 151
 Andrianov A., 43
 Andriyanov N., 127, 137
 Anikin A., 43
 Antsiperov V., 83, 191
 Apishev M., 149
 Arbel J., 21
- B**
- Bakhteev O., 29, 37
 Baklanov A., 107
 Beklaryan A., 81
 Beklaryan L., 81
 Berezin A., 119
 Berikov V., 11
 Bondur V., 123
 Boyko A., 97, 157
 Brazovskii K., 161
 Buchonov S., 165
 Bugaev A., 191
- C**
- Chelichkov A., 25
 Chernousova E., 103
- D**
- Dementiev V., 137
 Djukova E., 13
 Dvoenko S., 17
 Dvurechensky P., 79
- E**
- Efimov A., 167
- Efimov Y., 129, 139, 143
 Emelyanov G., 145
- F**
- Falomkina O., 25
 Fedotov N., 135
 Fedotova S., 115
 Fetisova N., 89
 Filipenkov N., 85
 Filippov S., 171
 Florinsky I., 171
- G**
- Gabova A., 159
 Gasnikov A., 77, 79
 Geppener V., 91
 Gerget O., 163
 Gimadi E., 45, 47, 49, 51
 Gneushev A., 109
 Goncharov E., 51
 Gorbatsevich V., 173
 Gornov A., 43, 53, 75
 Gorokhovskiy K., 125
 Grabovoy A., 37
 Gracheva I., 111
- I**
- Ianina A., 147
 Ignatiev V., 125
 Isachenko R., 87
 Istomin A., 49
 Ivanova A., 79
 Ivanova E., 119
- K**
- Karabanov A., 159

- Karatsuba E., 185
 Karpov L., 187
 Karpov Yu., 187
 Kelmanov A., . 55, 57, 59, 61,
 63, 65, 67
 Khachay D., 71
 Khachay M., ..69, 71, 73, 107
 Khairtdinov M., 175
 Khamidullin S., ... 59, 61, 63
 Khandeev V., . 55, 57, 59, 61,
 63, 65
 Khanykov I., 113
 Khaydarov A.,173
 Kirilyuk I., 99
 Komarov D., 131
 Komuro T.,169
 Kopylov A.,111, 141
 Korobkin M., 129, 139
 Kovalevsky V., 175
 Kozlov A.,145
 Krasotkina O., 101, 103, 105,
 165
 Kulikova L., 167
 Kurochkina A., 45, 47
 Kushnir O.,115
- L**
- Lange A.,153
 Lange M., 153
 Lebedev M., 131
 Lemtyuzhnikova D., 117
- M**
- Makarova A., 39
 Malkovsky N., 77
 Mandrikova B., 91
 Mandrikova O., 89, 91
 Manilo L., 93
- Mansurov G., 191
 Masliakov G., 13
 Matveev I., 129, 133, 139, 143
 Medvedev A., 105
 Meldo A.A.,33, 35
 Mikhailova L., 67
 Mikhaylov D., 145
 Mishin D., 51
 Moiseev A., 135
 Molchanova D., 97
 Moretti P., 185
 Morozov A., ...101, 105, 159,
 191
 Mottl V., ..101, 103, 105, 165
 Murashov D., 119
 Murynin A., 123, 125
- N**
- Nagornaya E., 45
 Naumov V., 179
 Nedelko V., 19
 Nédich A.,77
 Nelyubina E.,179
 Neychev R., 27
 Neznakhina K., 73
 Nikolsky A., 169
 Nosova S., 31
 Novik V., 133
 Novikov I.,11
- O**
- Odinokikh G., ..129, 139, 143
 Ogorodnikov I., 69
 Oplachko E., 183
- P**
- Panasenko A.,55, 57
 Pankratov A., 95

- Pankratova N., 97
 Paprotskiy S., 191
 Pasyukov M., 107
 Pekker J., 161
 Petrova M., 85
 Petrova N., 191
 Polozov Yu., 89
 Polupanov A., 191
 Prokofyev P., 13
 Pshenichny D., 17
 Pyatkin A., 59, 61, 63, 65
 Pyatkov M., 95
 Pytyev Y., 25
- R**
- Rakova K., 125
 Rodionov D., 141
 Rogozin A., 77
 Romanchenko S., 67
 Ryazanov V., 179
 Rybka E., 103
 Rykunov S., 97, 157, 183
 Rykunova E., 157
- S**
- Samsonov N., 109
 Selezneva E., 171
 Senko O., 99
 Seredin O., 115, 141
 Shamardin Y., 45, 63
 Shenmaier V., 63
 Shevyakov A., 45
 Shin E., 49
 Shiskin S., 25
 Smetanin Yu., 187
 Solomatin I., ... 129, 139, 143
 Strijov V., 27, 29, 37, 87
 Sulimova V., 39, 165
- Sushkova O., 159, 191
 Syemov A., 135
- T**
- Tetuev R., 95
 Tikhonov D., 167
 Tsidulko O., 47
 Turlapov V., 31, 169
- U**
- Uribe C., 77
 Ustinin M., 97, 157, 183
 Utkin L.V., 33, 35
- V**
- Vasiliev E., 169
 Vasiliev K., 127, 137
 Vinogradov A., 179
 Vizilter Y., 131, 173
 Vladimirova M., 21
 Vorontsov K., 147, 149
 Voskoboinikova G., 175
 Vygolov O., 131
- W**
- Windridge D., 165
- Y**
- Yakovlev A., 173
 Yakovlev P., 41
 Yankovskaya A., 161, 163
 Yanushko A., 191
- Z**
- Zalyaev T., 91
 Zarayskaya Y., 171
 Zarodnyuk T., 75
 Zasukhin S., 181
 Zasukhina E., 181
 Zharikov I., 149
 Zubyuk A., 23

MachineLearning.ru

<http://www.machinelearning.ru/>

Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. Цели ресурса — сконцентрировать информацию о достижениях ведущих научных школ; способствовать обмену опытом, накоплению и распространению научных знаний; предоставить площадку для виртуальных научных семинаров и обсуждений.

Журнал «Машинное обучение и анализ данных»

<http://jmla.org>

Журнал Машинное обучение и анализ данных публикует новые теоретические и обзорные статьи с результатами научных исследований в области искусственного интеллекта, теоретической информатики и приложений. Цель журнала — развитие теории машинного обучения, интеллектуального анализа данных и методов проведения вычислительных экспериментов. Принимаются статьи на русском и английском языках.

Научное издание

ИНТЕЛЛЕКТУАЛИЗАЦИЯ
ОБРАБОТКИ ИНФОРМАЦИИ
ИОИ-2018

Тезисы докладов
12-й Международной конференции
(Москва, Россия – Гаэта, Италия)

Напечатано с готового оригинал-макета

Сдано в набор 01.09.18. Подписано в печать 26.09.18.
Формат 60×90/16. Бумага офсетная. Печать цифровая.
Усл.-печ. л. 13,63. Уч.-изд. л. 10,0. Тираж 300 экз.

Заказ № 1043.

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, офис 101

ISBN 978-5-94588-236-2

