

РОССИЙСКАЯ АКАДЕМИЯ НАУК
ОТДЕЛЕНИЕ МАТЕМАТИЧЕСКИХ НАУК РАН
ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР ИМ. А. А. ДОРОДНИЦЫНА РАН

ТАВРИЧЕСКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ
ИМ. В. И. ВЕРНАДСКОГО

НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК БЕЛАРУСИ

ПРИ ПОДДЕРЖКЕ

РОССИЙСКОГО ФОНДА ФУНДАМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ
КОМПАНИЙ ФОРЕКСИС И ЦСПИР

Интеллектуализация обработки информации

ИОИ-10

Греция, о. Крит

Тезисы докладов
10-й международной конференции

ТОРУС
ПРЕСС  МОСКВА
2014

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

И 73

ББК 22.1:32.973.26-018.2

Интеллектуализация обработки информации:

И 73 10-я международная конференция. Греция, о. Крит, 4–11 октября 2014 г.: Тезисы докладов. — М.: Торус Пресс, 2014. — 266 с.

ISBN 978-5-94588-154-9

В сборнике представлены тезисы докладов 10-й Международной конференции «Интеллектуализация обработки информации–2014», проводимой Вычислительным центром им. А. А. Дородницына РАН, Таврическим национальным университетом им. В. И. Вернадского и Национальной академией наук Беларуси при финансовой и организационной поддержке РФФИ и компаний Форексис и ЦСПиР.

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.

ББК 22.1:32.973.26-018.2

ISBN 978-5-94588-154-9

© Авторы докладов, 2014

© ВЦ РАН, ТНУ, 2014

Международный оргкомитет

Сопредседатели: Журавлев Юрий Иванович,
академик РАН, Россия
Сергиенко Иван Васильевич,
академик НАН Украины
Абламейко Сергей Владимирович,
академик НАН Беларуси

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

Борисова Татьяна Игоревна
Громов Андрей Николаевич
Дударенко Марина Алексеевна
Ивахненко Андрей Александрович, *к.ф.-м.н.*
Ивкин Никита Петрович
Инякин Андрей Сергеевич, *к.ф.-м.н.*
Помазкова Евгения Владимировна
Татарчук Александр Игоревич, *к.ф.-м.н.*

Программный комитет

Сопредседатели: Рудаков Константин Владимирович,
чл.-корр. РАН, Россия
Донской Владимир Иосифович, *д.ф.-м.н., Россия*
Зорин Денис Николаевич, *профессор, США*

Учёный секретарь: Стрижов Вадим Викторович, *к.ф.-м.н.*

Сойфер Виктор Александрович, *чл.-корр. РАН*
Визильтер Юрий Валентинович, *д.ф.-м.н.*
Воронцов Константин Вячеславович, *д.ф.-м.н.*
Громова Ольга Алексеевна, *д.м.н.*
Гупал Анатолий Михайлович, *д.ф.-м.н.*
Двоенко Сергей Данилович, *д.ф.-м.н.*
Краснопрошин Виктор Владимирович, *д.т.н.*
Местецкий Леонид Мойсеевич, *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.*
Тузиков Александр Васильевич, *д.ф.-м.н.*
Устинин Михаил Николаевич, *д.ф.-м.н.*
Чуличков Алексей Иванович, *д.ф.-м.н.*
Хачай Михаил Юрьевич, *д.ф.-м.н.*
Шананин Александр Алексеевич, *д.ф.-м.н.*

**Российская секция
10-й Международной конференции
«Интеллектуализация обработки информации»**

Оргкомитет

Сопредседатели: Журавлев Юрий Иванович, *академик РАН*
Учёный секретарь: Чехович Юрий Викторович, *к.ф.-м.н.*

Борисова Татьяна Игоревна
Громов Андрей Николаевич
Дударенко Марина Алексеевна
Ивахненко Андрей Александрович, *к.ф.-м.н.*
Ивкин Никита Петрович
Инякин Андрей Сергеевич, *к.ф.-м.н.*
Помазкова Евгения Владимировна
Татарчук Александр Игоревич, *к.ф.-м.н.*

Программный комитет

Сопредседатели: Рудаков Константин Владимирович, *чл.-корр. РАН*
Донской Владимир Иосифович, *д.ф.-м.н.*
Учёный секретарь: Стрижов Вадим Викторович, *к.ф.-м.н.*

Сойфер Виктор Александрович, *чл.-корр. РАН*
Визильтер Юрий Валентинович, *д.ф.-м.н.*
Воронцов Константин Вячеславович, *д.ф.-м.н.*
Громова Ольга Алексеевна, *д.м.н.*
Двоенко Сергей Данилович, *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.*
Устинин Михаил Николаевич, *д.ф.-м.н.*
Чуличков Алексей Иванович, *д.ф.-м.н.*
Хачай Михаил Юрьевич, *д.ф.-м.н.*
Шананин Александр Алексеевич, *д.ф.-м.н.*

Рецензенты

Андиперов В. Е.
Бахтеев О. Ю.
Вальков А. С.
Визильтер Ю. В.
Губочкин И. В.
Двоенко С. Д.
Дьяконов А. Г.
Загоруйко Н. Г.
Ивахненко А. А.
Игнатов Д. И.
Инякин А. С.
Каркищенко А. Н.
Козлов И. В.
Копылов А. В.
Красоткина О. В.
Кудинов М. С.
Кузнецов М. П.
Кузьмин А. А.

Кушнир О. А.
Ланге М. М.
Лукашевич Н. В.
Майсурадзе А. И.
Максимов Ю. В.
Матвеев И. А.
Матросов М. П.
Миркин Б. Г.
Митекин В. А.
Мнухин В. Б.
Мурашов Д. М.
Неделько В. М.
Панов А. И.
Панов М. Е.
Полежаев В. А.
Потапенко А. А.
Рейер И. А.
Рудой Г. И.

Сафонов И. В.
Сенько О. В.
Середин О. С.
Сотнезов Р. М.
Стенина М. М.
Стрижов В. В.
Сулимова В. В.
Таханов Р. С.
Торшин И. Ю.
Турдаков Д. Ю.
Хачай М. Ю.
Хританков А. С.
Царьков С. В.
Черепанов Е. В.
Чичева М. А.
Чуличков А. И.
Шелепов В. Ю.

Краткое оглавление

Теория классификации	8
Метрические методы анализа данных	34
Дискретно-логические методы классификации	46
Анализ временных рядов	58
Сложность вычислений и оптимизация	70
Обработка изображений	104
Анализ изображений	112
Анализ видеопоследовательностей	122
Распознавание изображений	138
Анализ формы изображений	146
Приложения: медицина	158
Приложения: биоинформатика	176
Приложения: экономические системы	182
Приложения: обработка естественного языка	194
Приложения: геофизика	216
Приложения: информационная безопасность	228
Содержание	234
Авторский указатель	257

Brief contents

Theory of Classification	8
Distance-Based Data Analysis	34
Rule-Based Classification	46
Time Series Analysis	58
Computational Complexity and Optimization	70
Image Processing	104
Image Analysis	112
Video Analysis	122
Image Recognition	138
Images Shape Analysis	146
Applications: Medicine	158
Applications: Bioinformatics	176
Applications: Economics	182
Applications: Natural Language Processing	194
Applications: Geophysics	216
Applications: Information Security	228
Contents	234
Author index	261

Обучение распознаванию образов в произвольных метрических пространствах

Абрамов Вадим Игоревич^{1*}

v.abramov.work@gmail.com

*Середин Олег Сергеевич*²

oseredin@yandex.ru

*Моттль Вадим Вячеславович*³

vmottl@yandex.ru

¹Москва, Московский физико-технический институт

²Тула, Тульский государственный университет

³Москва, Вычислительный центр РАН

Основная математическая идея линейного подхода к восстановлению зависимостей заключается в погружении произвольного множества объектов реального мира, рассматриваемого как метрическое пространство, в некоторое расширенное линейное пространство со скалярным произведением. Эту идею удастся полностью реализовать, только если исходная метрика является *евклидовой*, т. е. обладает некоторыми особенно удобными свойствами [1]. Инструментом моделирования искомой зависимости в этом случае является поиск в расширенном линейном пространстве так называемого диполя — пары элементов, позволяющих характеризовать всякий объект реального мира его расстоянием до соответствующей дискриминантной гиперплоскости — множества элементов, равноотстоящих от узлов диполя. Однако для многих практических задач анализа данных естественная исходная метрика оказывается принципиально неевклидовой. Тогда исходное метрическое пространство удастся вложить лишь в *псевдоевклидово* линейное пространство, в котором метрика определена не для всех пар элементов и разрушается понятие дискриминантной гиперплоскости. В докладе рассматриваются пути преодоления этой трудности, основанные на том факте, что для элементов метрического пространства объектов реального мира, погружаемого в псевдоевклидово линейное пространство как подмножество изолированных точек, сохраняется корректное значение метрики.

Работа поддержана грантом РФФИ № 14-07-00661-а.

- [1] *Абрамов В. И., Середин О. С., Моттль В. В.* Обучение распознаванию образов в евклидовых метрических пространствах по методу опорных объектов // Известия ТулГУ, Естественные науки. Тула: Издательство ТулГУ, 2013. Вып. 2. С. 119–136.

Pattern recognition in arbitrary metric spaces

*Abramov Vadim*¹★

v.abramov.work@gmail.com

*Seredin Oleg*²

oseredin@yandex.ru

*Mottl Vadim*³

vmottl@yandex.ru

¹Moscow, Moscow Institute of Physics and Technology

²Tula, Tula State University

³Moscow, Computing Center of the RAS

The main mathematical idea of the linear approach to dependence estimation is embedding the given set of real-world objects, which is treated as a metric space, into an extended linear space with inner product. This idea lends itself to a complete realization only if the original metric is the *Euclidean* one, i. e., possesses some especially useful features [1]. In this case, the instrument of estimating the hidden dependence is determination of the so-called dipole in the extended linear space — an ordered pair of elements that allows to characterize any real-world object by its distance from the respective discriminant hyperplane, i. e., the set of elements equidistant from the nodes of the dipole. However, in many practical problems of data analysis, the natural original metric is not Euclidean in principle. Then, the original metric space can be embedded only into a *pseudo-Euclidean* linear space, in which the metric is defined not for all pairs of elements and, thus, the notion of discriminant hyperplane gets destroyed. In this paper, the authors consider the ways of overcoming this difficulty, which are underlied by the fact that the metric space of real-world objects is embedded into the pseudo-Euclidean linear space as a subset of isolated points, and the correct values of the original metric are preserved for them.

This research is funded by RFBR, grant 14-07-00661-a.

- [1] Abramov V.I., Seredin O.S., Mottl V.V. 2013. Obuchenie raspoznavaniyu obrazov v evklidovykh metrisheskikh prostranstvakh po metodu opornykh ob"ektov [Pattern recognition in Euclidean metric spaces by support vector method]. *Izvestiya of Tula State University, Natural Sciences* 2:119–136. [In Russian.]

Об одной модификации алгоритмов вычисления оценок

Журавлёв Юрий Иванович

zhur@ccas.ru

Москва, Вычислительный центр им. Дородницына Российской академии наук

В классе алгоритмов вычисления оценок [1] наиболее сложными для определения являются системы опорных множеств и параметры, определяющие наличие близости в множестве значений признаков. В работе разработаны методы, позволяющие заменить эти компоненты на решение задач таксономии с помощью простых алгоритмов. Для определения остальных компонент алгоритма предложены легко реализуемые отображения таблиц обучения и описаний распознаваемых объектов на два подмножества.

Пусть задана таблица обучения $(a_{i1}, \dots, a_{ik}, a_{ik+1}, \dots, a_{in})$, $i = 1, \dots, m$, классы K_j , $j = 1, \dots, l$, и распознаваемый объект $(b_1, \dots, b_n) = \hat{S}$. Формируем бинарные матрицы для классов K_j и распознаваемого объекта \hat{S} . Обозначим через B_k , $k = 1, \dots, n$, алгоритмы, вычисляющие близость $B_k(b_k, a_{ik})$, равную 1 (близость есть) или 0 (близость отсутствует). Формируем для объектов $S_i = (a_{i1}, \dots, a_{ik}, \dots, a_{in})$ и $\hat{S} = (b_1, \dots, b_k, \dots, b_n)$ строку бинарной матрицы:

$$S'(S_i, \hat{S}, K_j) = (c_{i1}, \dots, c_{ik}, \dots, c_{in}),$$

$$c_{ik} = \begin{cases} 1, & S_i \in K_j, B_k(b_k, a_{ik}) = 1 \text{ или } S_i \notin K_j, B_k(b_k, a_{ik}) = 0; \\ 0, & S_i \notin K_j, B_k(b_k, a_{ik}) = 1 \text{ или } S_i \in K_j, B_k(b_k, a_{ik}) = 0. \end{cases}$$

Полученные таким образом строки S' делим на подмножества $M_1: S_i \in K_j$, $M_2: S_i \notin K_j$ и получаем бинарную матрицу, строки которой можно считать единицами или нулями не всюду определенной булевой функции.

Задача распознавания и задачи вычисления «весомых» параметров алгоритмов вычисления оценок решаются далее логическими алгоритмами или алгоритмами линейной или билинейной оптимизации.

- [1] *Журавлёв Ю. И.* Об алгебраическом подходе к решению задач распознавания или классификации // Проблемы кибернетики, 1978. Т. 33.

On a new modification of estimation algorithms

Zhuravlev Yuri

zhur@ccas.ru

Moscow, Dorodnicyn Computing Center of RAS

The main complexity of estimation algorithms [1] is to define systems of support sets and parameters, which define the proximity in the range of features. The methods that allow to substitute this components with solving the problems of taxonomy with elementary algorithms are elaborated. To obtain the rest of the components of the algorithm, some easily applicable mappings of learning matrices and information vectors of new objects into two subsets are proposed.

Consider a learning matrix $(a_{i1}, \dots, a_{ik}, \dots, a_{in})$, $i = 1, \dots, m$, the classes K_j , $j = 1, \dots, l$, and a new object $\hat{S} = (b_1, \dots, b_n)$ to be recognized. Let produce binary matrices for classes K_j and a new object \hat{S} . Let use notation B_k , $k = 1, \dots, n$, for the algorithms that compute proximity $B_k(b_k, a_{ik})$ which is equal to 1 (when there is proximity) or 0 (when there is no proximity). For objects $S_i = (a_{i1}, \dots, a_{ik}, \dots, a_{in})$ and $\hat{S} = (b_1, \dots, b_k, \dots, b_n)$, let form a row of a binary matrix

$$S'(S_i, \hat{S}, K_j) = (c_{i1}, \dots, c_{ik}, \dots, c_{in})$$

where

$$c_{ik} = \begin{cases} 1, & S_i \in K_j, B_k(b_k, a_{ik}) = 1 \text{ or } S_i \notin K_j, B_k(b_k, a_{ik}) = 0; \\ 0, & S_i \notin K_j, B_k(b_k, a_{ik}) = 1 \text{ or } S_i \in K_j, B_k(b_k, a_{ik}) = 0. \end{cases}$$

Let divide the rows S' obtained this way into subsets $M_1: S_i \in K_j$, $M_2: S_i \notin K_j$ and thus obtain a binary matrix with the rows that can be treated as ones and zeros of partially defined boolean function.

The recognition problem and the problem of computing “informative” parameters for estimation algorithms are solved onwards with logic algorithms or algorithms of linear or bilinear optimization.

- [1] Zhuravlev Y. I. 1978. Ob algebraicheskom podhode k resheniyu zadach raspoznavaniya ili klassifikatsii [On algebraic approach to recognition and classification problems]. *Problemy kibernetiki* [Problems of cybernetics] 33.

Обучение распознаванию без переобучения

Загоруйко Николай Григорьевич^{1,2} zag@math.nsc.ru

Кутненко Ольга Андреевна^{1,2*} olga@math.nsc.ru

*Зырянов Александр Олегович*² alexander.zyryanov44@gmail.com

*Леванов Дмитрий Александрович*¹ levanovd@gmail.com

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

²Новосибирск, Новосибирский государственный университет

Проблема борьбы с переобучением алгоритмов распознавания является одной из центральных проблем анализа данных. В работе описан алгоритм, который выбирает подмножество наиболее информативных объектов и признаков, и останавливает процесс обучения в точке, в которой начинается переобучение.

Для вычисления сходства между объектами используется тернарная относительная мера — функция конкурентного сходства (FRiS-функция), оценивающая сходство объекта z с a в конкуренции с b :

$$F(z, a|b) = \frac{r(z, b) - r(z, a)}{r(z, b) + r(z, a)}.$$

В фиксированном признаковом пространстве обучение состоит в последовательном наращивании количества эталонных объектов («столпов»). На каждом шаге увеличения числа столпов вычисляется качество описания выборки (или оценка разделимости классов), которая зависит от сходства объектов с эталонами и от числа эталонов. Обнаружено, что в точке перегиба кривой, описывающей разделимость классов, заканчивается обучение и начинается переобучение. Выбор информативных признаков делается алгоритмом FRiS-Grad.

В работе предлагается метод принятия решений с учетом взвешенных расстояний. Описываются результаты тестирования алгоритма на модельных задачах [1].

Работа поддержана грантом РФФИ № 14-01-00039.

- [1] *Загоруйко Н. Г., Кутненко О. А., Зырянов А. О., Леванов Д. А.* Обучение распознаванию без переобучения // Машинное обучение и анализ данных, 2014. Т. 1. № 7. С. 891–901 <http://jmla.org/papers/doc/2014/no7/Zagoruiko2014Learning.pdf>.

Learning to recognition without overfitting

Zagoruiko Nikolay^{1,2}

zag@math.nsc.ru

Kutnenko Olga^{1,2*}

olga@math.nsc.ru

Zyryanov Alexander²

alexander.zyryanov44@gmail.com

Levanov Dmitriy¹

levanovd@gmail.com

¹Novosibirsk, Sobolev Institute of Mathematics, SB RAS

²Novosibirsk, Novosibirsk State University

The problem of dealing with overfitting of recognition algorithms is one of the central problems in data mining. The algorithm that selects the most informative subset of objects and features and stops the learning process at the point at which overfitting starts is described.

To calculate the similarity between objects the ternary relative measure called the function of rival similarity (FRiS-function) is used. It evaluates the similarity between object z and a vs. b :

$$F(z, a|b) = \frac{r(z, b) - r(z, a)}{r(z, b) + r(z, a)}.$$

Education includes the consistent increase of number of reference objects (stolps) in a fixed feature space.

On each step of augmentation of number of stolps, the estimation of quality of the description of sample (or divisibility of classes) is calculated. It depends on similarity of objects to stolps and quantity of stolps. It was found that at the inflection point of the curve describing the separability of classes, training finishes and overfitting starts. Selection of informative features is made by the algorithm FRiS-Grad.

This paper proposes a decision-making method based on weighted distances. The results of testing of the algorithm on simulated and real problems are described [1].

This research is funded by RFBR, grant 14-01-00039.

- [1] Zagoruiko N.G., Kutnenko O.A., Zyryanov A.O., Levanov D.A. 2014. Learning to recognition without overfitting. *Machine Learning and Data Analysis* 1(7):891–901 <http://jmla.org/papers/doc/2014/no7/Zagoruiko2014Learning.pdf>.

Экспериментальное исследование метода неявной кросс-валидации в задаче беспризнакового обучения распознаванию образов

*Левдик Павел Владимирович*¹* cold62@mail.ru
*Черноусова Елена Олеговна*¹ lena-ezhova@rambler.ru
*Сунгуров Дмитрий Сергеевич*¹ elekt@bk.ru
*Моттль Вадим Вячеславович*² vmottl@yandex.ru

¹Москва, Московский физико-технический институт

²Москва, Вычислительный центр РАН

В отличие от обычной кросс-валидации, требующей многократного повторения процесса обучения при разных разбиениях обучающей совокупности, метод неявной кросс-валидации предполагает однократное вычисление критерия, зависящего только от исходных данных и варьируемых параметров регуляризации. В данном докладе экспериментально сравниваются методы обычной и неявной кросс-валидации в задаче распознавания образов на основе представления объекта вектором вторичных признаков, образованных сравнением его сходства с каждым из объектов обучающей совокупности. Отбор подмножества релевантных обучающих объектов, с которыми достаточно сравнивать предъявленный объект, обеспечивает критерий обучения Elastic Net SVM (support vector machine), содержащий структурный параметр, определяющий степень селективности отбора. В качестве источника экспериментальных данных использован известный массив «Chicken Pieces Silhouettes Database», содержащий 446 силуэтных изображений частей курицы (крылышко, спинка, ножка, бедрашко, грудка), представленных 44 разными функциями парного сравнения изображений. Оказалось, что неявная кросс-валидация выделяет всего несколько релевантных объектов, достаточных для решающего правила, в то время как обычный скользящий контроль в данной задаче оказался нечувствительным к селективности отбора [1].

Работа поддержана грантом РФФИ № 14-07-00661-а.

- [1] *E. Chernousova, P. Levdik, A. Tatarchuk, V. Mottl, D. Windridge.* Non-enumerative cross validation for the determination of structural parameters in feature-selective SVMs // 22th Conference (International) on Pattern Recognition Proceedings. Stockholm, Sweden, 2014.

Experimental study of hypothetical cross validation applied to determination of structural parameters in featureless pattern recognition

Levdik Pavel^{1*}

cold62@mail.ru

*Chernousova Elena*¹

lena-ezhova@rambler.ru

*Sungurov Dmitry*¹

elekt@bk.ru

*Mottl Vadim*²

vmottl@yandex.ru

¹Moscow, Moscow Institute of Physics and Technology

²Moscow, Computing Center of RAS

As distinct from the usual cross validation, which requires multiple repetition of the training process with different partitionings of the training set, the method of hypothetical cross validation implies computation of a single criterion that depends only on the original data set and regularization parameters to be varied. In this paper, the methods of usual and hypothetical cross validation are experimentally compared in the problem of relational pattern recognition, which is based on the representation of any real-world object by the vector of secondary features formed by measuring its proximity to each of the objects in the training set. The selection of a subset of relevant training objects, which would be sufficient for the representation of the given object, is provided by the training criterion Elastic Net that includes a structural parameter responsible for the selectivity level. As the source of experimental data, the well-known data set “Chicken Pieces Silhouettes Database” is used, which contains 446 silhouette images of specific chicken parts (wing, back, drumstick, thigh, breast) represented by 44 different pairwise image comparison functions. The experiments have shown that the hypothetical cross validation selects only a few number of relevant training objects that are completely sufficient for the classification rule, whereas the usual leave-one-out procedure, which requires, in addition, a much greater amount of computation, is insensitive to the selectivity level [1].

This research is funded by RFBR, grant 14-07-00661-a.

- [1] E. Chernousova, P. Levdik, A. Tatarchuk, V. Mottl, D. Windridge. 2014. Non-enumerative cross validation for the determination of structural parameters in feature-selective SVMs. *22th Conference (International) on Pattern Recognition Proceedings*. Stockholm, Sweden.

Сравнение искаженных гистограмм вероятностными методами

Лепский Александр Евгеньевич

`alex.lepskiy@gmail.com`

Москва, Национальный исследовательский университет «Высшая школа экономики»

Проблема сравнения гистограмм возникает при решении многих прикладных задач анализа данных. Под сравнением будем понимать определение на множестве гистограмм вида $U = (x_i, u_i)_{i \in I}$, $x_i < x_{i+1}$, $i \in I$, отношения типа «больше-меньше». Примерами таких задач являются: сравнение результатов различных опытов, сравнение показателей функционирования однородных систем, принятие решений в условиях нечеткостной неопределенности, сравнение распределений доходов, ранжирование учащихся по результатам-гистограммам их оценок и т.д.

Для сравнения гистограмм в работе рассматриваются методы вероятностного подхода, в котором сравниваются некоторые числовые характеристики случайных величин, связанных со сравниваемой парой гистограмм. Необходимо учитывать, что сами гистограммы могут быть заданы с той или иной степенью неточности. Характер и причины этих неточностей могут быть различны. В работе под искажением понимаются интервальные поточечные некоррелированные изменения. Возникает задача определения, может ли данное искажение гистограмм изменить их сравнение определенным методом на противоположное.

В работе получены необходимые и достаточные условия на искажения пары гистограмм, гарантирующие сохранение знака сравнения. Исследование проведено для трех вероятностных методов сравнения: математического ожидания, стохастического доминирования, стохастического предшествования. Утверждения проиллюстрированы исследованиями устойчивости сравнений гистограмм результатов ЕГЭ абитуриентов, поступивших в вузы.

Исследование осуществлено в рамках Программы фундаментальных исследований НИУ ВШЭ в 2014 году и при поддержке РФФИ, грант № 14-07-00189.

- [1] *Лепский А. Е.* Сравнение искаженных гистограмм вероятностными методами // Машинное обучение и анализ данных, 2014. Т. 1. №8. С. 949–965. <http://jmla.org/papers/doc/2014/no8/Lepskiy2014Comparison.pdf>.

Comparison of distorted histograms by probability methods

Lepskiy Alexander

alex.lepskiy@gmail.com

Moscow, Higher School of Economics

The comparison of histograms is necessary in many applied problems of data processing. The comparison of type “more-less” is considered in this paper. There are many approaches to such comparison. Some methods of probability approach were considered in this paper.

But the histograms may be distorted. We consider pointwise interval uncorrelated distortions of histograms. Then we have a problem to find the conditions on distortions under which the comparison of the two histograms is not changed.

This problem solved in the case of three popular probabilistic methods of comparison: comparison of mathematical expectations, comparison with help of principle of stochastic dominance, comparison with help of principle of stochastic precedence.

The characteristics of stability of histograms to distortion are introduced. These characteristics are calculated for histograms of USE (Unified State Exam) of applicants admitted in 2012 in Russian universities.

The study was implemented in the framework of The Basic Research Program of the Higher School of Economics. This work was supported by the grant 14-07-00189 of RFBR (Russian Foundation for Basic Research).

- [1] Lepskiy A. 2014. Comparison of distorted histograms by probability methods. *Machine Learning and Data Mining* 1(7):949–965. <http://jmla.org/papers/doc/2014/no8/Lepskiy2014Comparison.pdf>.

Обучение монотонного классификатора ближайшего соседа по немонотонной выборке

Махина Галина Анатольевна

galkabird@gmail.com

Румыния, Бухарест, ModuleWorks

Монотонные классификаторы находят применение в прикладных задачах с ограничениями монотонности (когда известно, что чем больше значение признака, тем выше вероятность класса 1), а также при построении композиций классификаторов.

Построение монотонной функции по заданной выборке обычно предполагает решение двух подзадач. Сначала производится монотонизация исходной выборки путём исключения или переклассификации объектов, нарушающих монотонность. Затем строится монотонная функция, проходящая через точки монотонизированной выборки.

Предлагается монотонный классификатор ближайшего соседа (MNN), который не требует трудоёмкого предварительного этапа монотонизации выборки. Введение специальной функции M -расстояния между классифицируемым объектом и объектами обучающей выборки гарантирует, что классификатор ближайшего соседа является монотонной функцией при любой выборке. Аналогичная конструкция была предложена ранее в [1], но применялась только к монотонизированным выборкам, и только для задачи восстановления частично заданных монотонных булевых функций. В докладе рассматривается её обобщение на случай произвольных дискретнозначных признаков.

Для предложенного алгоритма получены достижимые верхние и нижние оценки полного скользящего контроля (CCV), вычисляемые за полиномиальное время. Их минимизация позволяет производить отбор признаков в задачах классификации с ограничением монотонности.

- [1] *Махина Г. А.* О восстановлении монотонных булевых функций методом ближайшего соседа // 9-я Международная конференция «Интеллектуализация обработки информации», ИОИ-9. — М.: МАКС Пресс, 2012. — С. 78–81.

Learning isotonic nearest neighbor classifier from non-isotonic training set

Makhina Galina

`galkabird@gmail.com`

Romania, Bucharest, ModuleWorks

Isotonic classifiers are used in applications with monotonicity constraint when it is known beforehand that the probability of class 1 is increasing in value of features. Isotonic classifiers also can be used as aggregating functions for ensemble learning.

Learning isotonic classifier from a given training set usually involves two subproblems. At the first step, all objects violating a monotonicity constraint are eliminated or relabeled. The second step implies learning a classifier on monotone data.

We propose a monotone nearest neighbor (MNN) classifier which doesn't involve the first step of relabeling. The monotonicity of the classifier is ensured by introducing a special M -distance function to measure a distance between a training object and an object to classify. A similar technique was applied in [1], though that time it was considered only for monotone data set and only for Boolean-valued classifiers. Here we consider a generalized approach for the case of ordinal discrete-valued features.

For the MNN learning algorithm we induce the strict lower and upper bounds of the complete cross-validation (CCV) computable in polynomial time. Minimization of CCV allows for feature selection in classification problems with monotonicity constraint.

- [1] *Makhina G. A.* About reconstruction of monotone Boolean functions with nearest neighbor classifier // The 9th International Conference "intellectualization of Information Processing", Moscow, 2012, Pp. 78–81.

Некоторые вопросы оценивания качества методов построения решающих функций

Неделько Виктор Михайлович

nedelko@math.nsc.ru

Новосибирск, Институт математики СО РАН

Работа посвящена проблеме оценивания качества методов построения решающих функций в задачах классификации. Исследуется возможность использования подхода, основанного на задании эталонного набора тестовых задач, а также связь данной проблемы с проблемой оценивания риска в задаче классификации, в частности нахождения распределений, при которых погрешность оценки риска максимальна.

В качестве тестовых единиц, на которых производится сравнение методов, предлагается использовать специальным образом подобранные классы распределений. Классы распределений подбираются таким образом, чтобы статистическое моделирование на них по возможности полно отражало особенности тестируемого метода обучения. Это, в частности, означает, что класс распределений должен являться параметрическим семейством, один из параметров которого есть наименьшее достигаемое значение риска в заданном классе решающих правил.

Другим важным параметром является величина смещения эмпирического риска. Установлено, что для гистограммного классификатора «наихудшее» распределение (для которого смещение эмпирического риска максимально) является смесью равномерного (в пространстве переменных, используемых для прогноза) распределения и распределения, сосредоточенного в одной точке.

Подобного вида распределения легко могут быть заданы и в непрерывном пространстве переменных. Результаты статистического моделирования позволяют предположить, что на таких распределениях будет достигаться максимальное смещение эмпирического риска и для других (помимо гистограммного) методов классификации [1].

Работа поддержана грантом РФФИ № 14-01-00590.

- [1] *Неделько В. М.* Некоторые вопросы оценивания качества методов построения решающих функций // Вестник Томского гос. ун-та. Управление, вычислительная техника и информатика, 2013. Т. 3. № 24. С. 123–132.

Some aspects of estimating a quality of decision functions construction methods

Nedelko Victor

nedelko@math.nsc.ru

Novosibirsk, Institute of mathematics SB RAS

The paper is devoted to the problem of estimating a quality of decision functions construction methods in classification (pattern recognition) task. An approach based on constructing some special set of testing tasks is investigated.

Some kinds of parametrical sets of distributions are suggested for using as the testing tasks.

An important class of such distributions are comprised by the distributions delivering the maximal bias of empirical risk. In the case of histogram classifier, these distributions look as a mixture of uniform distribution and distribution collected at a point. The statistical modeling performed allows to evaluate an applicability of the results obtained [1].

This research is funded by RFBR, grant 14-01-00590.

- [1] Nedelko V. 2013. Some aspects of estimating a quality of decision functions construction methods. *Herald of Tomsk State University. J. Control and Computer Science* 3(24):123–132.

Обучение многоклассовому распознаванию образов по большим обучающим совокупностям

*Никишин Михаил Сергеевич*¹ j48sa150a.p3az@gmail.com
*Красоткина Ольга Вячеславовна*¹ o.v.krasotkina@yandex.ru
Моттль Вадим Вячеславович^{2*} vmottl@yandex.ru

¹Тула, Тульский государственный университет

²Москва, Вычислительный центр РАН

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

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

Работа поддержана грантом РФФИ № 14-07-00527-а.

- [1] Двоенко С.Д., Моттль В.В., Середин О.С. Процедура распознавания образов для случая многих классов, опирающаяся на совокупность функций степени достоверности для пар классов // Известия ТулГУ, Вычислительная техника, автоматика, управление, 1999. Т. 2. Вып. 2. С. 28–35.

Multiclass pattern recognition from large training sets

*Nikishin Michael*¹

j48sa150a.p3az@gmail.com

*Krasotkina Olga*¹

o.v.krasotkina@yandex.ru

*Mottl Vadim*² *

vmottl@yandex.ru

¹Tula, Tula State University

²Moscow, Computing Center of the RAS

The problem of multiclass supervised pattern recognition from large training sets is typical for many applications in engineering and scientific research. The existing training methods either suffer from too high computational complexity that hinders their application to large data sets or are in principle unable to work with more than two classes. Moreover, in many applications, it is required not only to indicate the most preferable class of the new object, but also to estimate the posterior probabilities of its class memberships over several classes.

The idea of this work consists in constructing a mathematical framework that harnesses the well-known probabilistic methodology of two-class pattern recognition, which yields the posterior probability of one class against the other. A mathematical framework of fusing several single two-class posterior probabilities into a unified posterior distribution over all the classes is suggested. This approach has displayed its ability to reduce the computational complexity of training to almost linear with respect to the size of the training set. Another advantage of the proposed approach is that it evidently lends itself to easy parallelization using a computer cluster, for instance, by way of loading the training set in parts associated with single pairs of classes [1].

This research is funded by RFBR, grant No. 14-07-00661-a.

- [1] Dvoenko S. D., Mottl V. V., Seredin O. S. 1999. A pattern recognition procedure for the case of several classes on the basis of the set of confidence functions for pairs of classes. *Izvestiya of Tula State University. Computers, Automation and Control*. 2(2):28–35. [In Russian.]

Комбинирование частичных порядков в задаче обучения по предпочтениям

*Стенина Мария Михайловна** mmedvednikova@gmail.com

Кузнецов Михаил Павлович mikhail.kuznecov@phystech.edu

Долгопрудный, Московский физико-технический институт
(государственный университет)

Решается задача обучения по предпочтениям с использованием частично упорядоченных множеств значений признаков. Задано множество объектов мощности m . Каждый объект описан набором признаков мощности n , причем на множестве значений каждого признака задано отношение частичного порядка. Другими словами, каждый j -й признак определяет j -е отношение попарного доминирования объектов. Требуется решить задачу построения оптимального отношения доминирования на множестве объектов, где оптимальность понимается в смысле близости построенного отношения объектов к истинному, заданному на обучающем подмножестве. В качестве частных случаев, определяемых свойствами отношения порядка на множестве объектов, рассматривается задача ранжирования и задача порядковой классификации с монотонными ограничениями.

Решение задачи проводится в два этапа. На первом этапе предлагается оценить матрицу попарного доминирования объектов путем комбинирования частичных порядков, соответствующих признакам. Значения элементов построенной матрицы принадлежат единичному отрезку действительной оси и выражают степень отношения доминирования соответствующих объектов. На втором этапе предлагается восстановить оптимальное отношение доминирования с использованием взвешенного графа, соответствующего построенной матрице. С помощью аппарата конусов, соответствующих частичным порядкам [1], доказываем, что построенная модель описывает достаточно широкое множество возможных решений. Предлагается метод сокращения размерности пространства параметров модели.

Работа поддержана грантом РФФИ № 14-07-31042.

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

Partial orders aggregation for preference learning problem

*Stenina Maria**

mmedvednikova@gmail.com

Kuznetsov Mikhail

mikhail.kuznecov@phystech.edu

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

A preference learning problem is solved using partial orders of feature sets. The problem is as follows The set of m objects is given. Each object is described by the set of n features such that every feature value set is a poset. In other words, every j th feature defines j th dominance relation between each pair of objects. The problem is to construct an optimal dominance relation on the set of objects, where optimality is defined as the proximity of the estimated dominance relation to the proper relation, given on the learning subset. In the special case of a preference learning problem, an ordinal ranking problem and an ordinal classification with monotonic constraints are considered. It is shown that different special problem settings correspond to the different properties of a partial relation.

A two-stage algorithm is suggested to solve a problem. On the first stage, a matrix of the pairwise dominance of objects is estimated by aggregating partial orders corresponding to the features. The values of the estimated matrix belong to the $[0, 1]$ segment and describe the rate of the corresponding objects dominance. On the second stage, it is proposed to estimate an optimal dominance relation using a weighted graph corresponding to the obtained matrix. Using the partial order cones technique [1], the present author proves that the constructed model describes a sufficiently wide set of possible solutions. A method of the parameter space dimension reduction is suggested.

This research is funded by RFBR, grant 14-07-31042.

- [1] Kuznetsov M., Strijov V. 2014. Methods of expert estimations concordance for integral quality estimation. *Expert Systems with Applications* 41(4):1988—1996.

Байесовские постановки задачи обучения распознаванию образов по методу опорных векторов с управляемой селективностью отбора информативных признаков объектов

*Татарчук Александр Игоревич**

`aitech@yandex.ru`

Моттль Вадим Вячеславович

`vmottl@yandex.ru`

Москва, Вычислительный центр РАН

Наиболее популярным методом обучения линейного классификатора по обучающей совокупности является метод опорных векторов (Support Vector Machine — SVM). Однако при всей эффективности метод опорных векторов остается эвристическим по своей конструкции. С момента его создания в мировой литературе был предпринят ряд попыток снабдить его некоторой вероятностной интерпретацией, однако эти интерпретации оставались незавершенными, не позволяющими в полной мере использовать вероятностный аппарат для наделения чрезвычайно популярного метода опорных векторов принципиально новыми свойствами. Предлагается специальная байесовская постановка обучения распознаванию двух классов объектов в линейном признаковом пространстве, приводящая к обобщению метода опорных векторов и являющаяся теоретической основой для создания новых методов обучения. Кроме того, стремление избежать переобучения заставляет строить селективные системы распознавания, основанные на сокращении множества признаков, характеризующих объекты распознавания. Предлагается использовать выбор априорного распределения направляющего вектора разделяющей гиперплоскости для наделения метода опорных векторов в его вероятностной интерпретации способностью автоматически отбирать подмножество информативных признаков с заданным уровнем селективности. На этой основе разработаны новые селективные модификации метода опорных векторов [1].

Работа поддержана грантами РФФИ 11-07-00409 и 14-07-00661.

- [1] *Seredin O., Mottl V., Tatarchuk A., Razin N., Windridge D.* Convex support and relevance vector machines for selective multimodal pattern recognition // 21th Conference (International) on Pattern Recognition Proceedings. — Tsukuba, Japan, 2012. P. 1647–1650.

Bayesian feature-selective formulations of the pattern recognition problem based on support vector machine

*Tatarchuk Alexander**

aitech@yandex.ru

Mottl Vadim

vmottl@yandex.ru

Moscow, Computing Center of the RAS

The Support Vector Machine (SVM) is the most popular method of finerring a linear classifier from the given training set. However, this method, despite its exceptional effectiveness, still remains a heuristic one by construction. Since the creation of the SVM method, several attempts have been endeavored to supply it with some probabilistic background, but all these interpretations were uncompleted and remained insufficient for endowing this extremely popular method with principally new properties. In this work, a special Bayesian formulation of the two-class pattern recognition problem in the linear feature space is proposed that results in an essential generalization of the SVM idea and serves as a theoretical basis for developing new training methods. Besides, the intent to avoid the overfitting danger leads to the necessity to create new feature-selective pattern recognition systems that would be based on reducing the set of object features. It is proposed to harness the choice of the *a priori* distribution of the discriminant hyperplane's direction vector as a mathematical instrument of providing the SVM, in its probabilistic formulation, with the ability to automatically select a subset of features with the preset selectivity level. The respective theoretical framework has resulted in developing new mathematically sound feature-selective modifications of the originally heuristic SVM idea [1].

The work was supported by RFBR, grants 11-07-00409 and 14-07-00661.

- [1] Seredin O., Mottl V., Tatarchuk A., Razin N., Windridge D. 2012. Convex support and relevance vector machines for selective multimodal pattern recognition. *21th Conference (International) on Pattern Recognition Proceedings*. Tsukuba, Japan. 1647–1650.

Отбор признаков в задаче обучения распознаванию образов при смещении решающего правила

*Турков Павел Анатольевич**

pavel.turkov@gmail.com

Красоткина Ольга Вячеславовна

o.v.krasotkina@yandex.ru

Тула, Тульский государственный университет

Рассматривается задача обучения распознаванию образов при смещении решающего правила, что означает наличие в генеральной совокупности процессов, влияние которых приводит к изменению зависимости между скрытой и наблюдаемой характеристиками объектов обучения. В задаче обучения с учителем единственным источником информации о состоянии генеральной совокупности является обучающая выборка, однако в случае «дрейфа» решающего правила речь уже идет не о выборке, а о потоке объектов обучения. Использование отбора признаков в таких задачах имеет как минимум два очевидных преимущества: во-первых, позволяет уменьшить объем обрабатываемых данных в условиях работы с потоком объектов обучения, во-вторых, уменьшает риск переобучения.

Использование механизма отбора признаков в методе обучения распознаванию образов при смещении решающего правила основано на байесовском подходе к оценке зависимостей. Для его применения введем в рассмотрение априорное распределение вектора параметров решающего правила, предполагая, что указанное распределение будет принадлежать классу экспоненциальных. Данный метод позволяет осуществлять отбор релевантных признаков непосредственно в процессе обучения, что дает возможность его применения в задачах распознавания на потоках данных [1].

Работа поддержана грантом РФФИ № 14-07-00964.

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

Feature selection in the pattern recognition problem under concept drift

*Turkov Pavel**

pavel.turkov@gmail.com

Krasotkina Olga

o.v.krasotkna@yandex.ru

Tula, Tula State University

The paper considers the pattern recognition problem under concept drift that means there are the processes in the universe which influence dependence between hidden and observable features of training instance. In the supervised learning problem, the single information source about universe state is a training set but in case of concept drift, we are speaking not about instance set but about instance stream. If only one instance enters every moment, such learning is incremental. If researcher receives some instances every moment, then batch learning occurs. Feature selection has two obvious advantages: first, it allows to decrease the size of processing data, and second, it decreases the chance of overfitting.

Using of feature selection mechanism in the pattern recognition problem under concept drift is based on Bayesian approach by the model estimation. For its application, *a priori* probability of decision rule parameters is considered and it is supposed that this probability belongs to the exponential class. This method allows to realize selection of relevant features in the learning process directly that makes it applicable in the pattern recognition problem with data stream [1].

This research is funded by RFBR, grant 14-07-00964.

- [1] Krasotkina O., Turkov P., Mottl V. 2013. Bayesian logistic regression in the pattern recognition problem under concept drift. *Izvestiya Tulskogo gos. un-ta. Tekhnicheskie nauki* 2:177–187.

Неявная кросс-валидация для выбора подмножества информативных признаков в задаче обучения распознаванию образов по методу опорных векторов

Черноусова Елена Олеговна^{1*}

lena-ezhova@rambler.ru

*Левдик Павел Владимирович*¹

cold62@mail.ru

*Моттль Вадим Вячеславович*²

vmottl@yandex.ru

¹Москва, Московский физико-технический институт

²Москва, Вычислительный центр РАН

Широким классом задач машинного обучения является восстановление зависимости по эмпирическим данным, среди которых наиболее типичной является задача классификации. Для задачи классификации общепринятым способом обучения является метод опорных векторов (SVM — support vector machine). Как правило, для упрощения модели требуется провести отбор признаков с целью оставить только самые информативные. Этого добиваются, в частности, путем включения в обучающий критерий SVM, наряду с традиционным квадратичным регуляризирующим членом, еще и модульного регуляризатора с некоторым дополнительным коэффициентом. Для того чтобы подобрать нужные значения теперь уже двух структурных параметров, отвечающих за соответствие рассматриваемой модели данным предположениям наблюдателя, обычно используют многократное повторение процесса обучения посредством кросс-валидации. Чтобы избежать перебора, в данном докладе предложен метод математической имитации кросс-валидации, используя естественные допущения о природе неизвестного вероятностного распределения, породившего обучающую выборку. Этот подход назван неявной кросс-валидацией. Таким образом, многократное повторение процесса обучения заменяется вычислением единственного критерия, зависящего только от исходных данных и параметров регуляризации [1].

Работа поддержана грантом РФФИ № 14-07-00661-а.

- [1] *E. Chernousova, P. Levdik, A. Tatarchuk, V. Mottl, D. Windridge.* Non-enumerative cross validation for the determination of structural parameters in feature-selective SVMs. 22th Conference (International) on Pattern Recognition Proceedings, 2014. Stockholm, Sweden.

Nonenumerative cross validation for the determination of structural parameters in feature-selective support vector machines

Chernousova Elena^{1*}

lena-ezhova@rambler.ru

*Levdik Pavel*¹

cold62@mail.ru

*Mottl Vadim*²

vmottl@yandex.ru

¹Moscow, Moscow Institute of Physics and Technology

²Moscow, Computing Center of the RAS

The wide class of machine-learning problems is dependency estimation and the particular case is classification. For solution classification problem, commonly used method is Support Vector Machine (SVM). Sometimes, for simplification of the model, it is necessary to make feature selection with an aim to save only the most informative ones, and the way to reach it is adding modular regularization term with some coefficient to learning criteria. Thus, learning criteria contains one or a few structural parameters, which answer for correspondence of the treated models with observer assumptions. In traditional approaches, multiple training repetitions would be required to determine the best values for these structural parameters via cross-validation. To avoid this, the authors seek for algorithmically emulate the cross-validation process using conservative assumptions as to the nature of the unknown probability distribution that produced the training set. This approach was called nonenumerative cross-validation. Thus, multiple training repetitions are exchanged with calculating the only criteria, which depends on only initial data and regularization parameters [1].

This research is funded by RFBR, grant 14-07-00661-a.

- [1] Chernousova E., Levdik P., Tatarchuk A., Mottl V., Windridge D. 2014. Non-enumerative cross validation for the determination of structural parameters in feature-selective SVMs. *22th Conference (International) on Pattern Recognition Proceedings*. Stockholm, Sweden.

Оценки, минимизирующие возможность потерь, и минимаксные оценки: сравнительный анализ

*Чуличков Алексей Иванович**

achulichkov@gmail.com

Юань Боюань

yuanboyuan1983@hotmail.com

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

Поставлена и решена задача оценивания значений функции $f : X \rightarrow R$ в заданных точках области X по результатам измерений конечного набора ее функционалов, выполненных с погрешностью. Оценивание основано на минимизации возможности потерь. Описана теоретико-возможностная модель измерений, в которой считается, что большие значения погрешности измерения каждого функционала менее возможны, чем малые. Дан критерий оценивания, минимизирующий возможность потерь.

Оценки минимальной возможности потерь сравниваются с минимаксными оценками значений функции, полученными при условии, что погрешность измерения каждого линейного функционала принадлежит заданному интервалу (т. е. для случая, когда значения ошибки измерений обладают равной возможностью внутри заданного интервала). Обсуждаются различия минимаксных оценок и оценок минимальной возможности потерь. Приведен пример оценивания параметров реального спектрометрического эксперимента [1].

Работа поддержана грантом РФФИ № 14-07-00409.

- [1] Чуличков А. И., Юань Б. Оценки, минимизирующие возможность потерь, и минимаксные оценки: сравнительный анализ // Машинное обучение и анализ данных, 2014 (в печати).

Estimation of minimum possibility of losses and minimax estimation: A comparative analysis

*Chulichkov Aleksey**

achulichkov@gmail.com

Yuan Boyuan

yuanboyuan1983@hotmail.com

Moscow, Moscow State University, Faculty of Physics

The problem of estimating the values of function $f : X \rightarrow R$ in the points $x_1, \dots, x_N \in X$ is solved. The estimations are based on the results of measurements of the finite set of linear functionals made with an error, and on minimization of the possible losses.

The paper describes the probabilistic measurement model in which it is believed that large values of measurement error of each linear functional are less possible than small. The criterion of estimation is to minimize the possibility of losses.

The estimates of the minimum possible losses are compared with the estimates of the function values that minimize their maximum error. Minimax estimates are obtained under the condition that the measurement error for each linear functional is in the given interval. In other words, the values of the measurement errors are of equal possibility into the specified interval. The differences between the minimax estimations and the estimations of minimum possible losses are discussed. An example of estimating the parameters of the real spectrometric experiment is given [1].

This research is funded by RFBR, grant 14-07-00409.

- [1] Chulichkov A., Yuan B. 2014 (in press). Estimation of minimum possibility of losses and minimax estimation: A comparative analysis. *Machine Learning and Data Analysis*.

Ансамбль алгоритмов кластерного анализа с весами

Бериков Владимир Борисович

berikov@math.nsc.ru

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

В работе решается задача кластерного анализа сложноструктурированных данных с использованием коллектива (ансамбля) различных алгоритмов. При этом предполагается, что каждый алгоритм «специализирован» (дает наилучшее качество решения) на определенном классе таких структур. Степень «компетентности» алгоритма определяется с помощью весов, определенных для каждой пары наблюдений. Оптимальные веса находятся на основе анализа наблюдаемых характеристик работы алгоритмов с использованием предложенной модели ансамбля.

В отличие от других существующих подходов предложенный метод назначает веса не только алгоритмам, но и парам различных объектов. Это дает возможность определять компетентность алгоритмов при классификации структур, к которым принадлежат эти пары. Для определения оптимальных весов используется критерий, основанный на найденном выражении для верхней границы вероятности непосредственно ненаблюдаемой ошибочной классификации. Эта ошибка оценивается по наблюдаемым характеристикам ансамбля.

Целесообразность применения предложенного метода подтверждена численным моделированием с помощью метода Монте-Карло. В проведенных экспериментах истинные метки классов сравнивались с полученным алгоритмом, что дает возможность объективно оценить качество решения в зависимости от числа «шумовых» переменных. Моделирование показало статистически значимое улучшение качества кластеризации для предложенного алгоритма по сравнению с аналогичным алгоритмом с равными весами. Эксперименты продемонстрировали эффективность коллективного подхода в кластерном анализе.

Работа поддержана грантом РФФИ №14-07-00249 и благотворительным фондом В. Потанина.

[1] *Berikov V.* Weighted ensemble of algorithms for complex data clustering // *Pattern Recogn. Lett*, 2014. Vol. 38. P. 99–106.

Ensemble of clustering algorithms using weights

Berikov Vladimir

berikov@math.nsc.ru

Novosibirsk, Sobolev Institute of Mathematics SB RAN

The paper considers a problem of clustering complex data composed from various structures. A collection of different algorithms is used for the analysis. The main idea is based on the assumption that each algorithm is “specialized” (as a rule, gives more accurate partition results) on particular types of structures. The degree of algorithm’s “competence” is determined by usage of weights attributed to each pair of observations. Optimal weights are specified by the analysis of partial ensemble solutions using the proposed model of clustering ensemble.

Unlike other existing approaches, the proposed one assigns weights not only to algorithms, but additionally to the pairs of objects. This allows one to consider the “competence” of algorithms in classifying different data structures. For finding optimal weights, the criterion based on the found expression for upper bound of error probability in partitioning an object pair into clusters was used. Directly unobserved error is associated with observable ensemble characteristics.

The applicability of the suggested approach in real clustering tasks is confirmed by numerical experiments with Monte-Carlo modeling. In these experiments, the true class labels were available that made it possible to assess the objective measure of algorithms’ performance for a given distribution family, in dependence on the number of noise features. Monte-Carlo simulations have demonstrated significant improvement of clustering accuracy for the proposed algorithm in comparison with equal weights averaging. The experiments have shown the benefits of the ensemble approach in cluster analysis [1].

This research is funded by RFBR, grant 14-07-00249 and by V. Potanin Foundation.

- [1] Berikov V. 2014. Weighted ensemble of algorithms for complex data clustering. *Pattern Recogn. Lett.* 38:99–106.

Двухкомпонентная функция качества кластеризации множества элементов, представленных парными сравнениями

Двоенко Сергей Данилович

`dsd@tsu.tula.ru`

Тула, Тульский государственный университет

В случае, когда отсутствует признаковое пространство, множество элементов представлено только результатами их парных сравнений в виде матрицы близостей или расстояний. Для корректной обработки таких данных необходимо модифицировать алгоритмы кластер-анализа и распознавания.

Рассмотрены варианты известного алгоритма k -средних, в которых не требуется вычислять собственно средние по кластерам. В новых версиях алгоритма k -средних выполняются перестановки на матрице парных сравнений так, что в случае помещения анализируемого множества объектов в признаковое пространство достигается тот же самый результат кластеризации.

Рассмотрена новая двухкомпонентная целевая функция качества кластеризации как минимизируемая комбинация внутрикластерных дисперсий (квадратов расстояний) с близостью кластеров между собой или, в двойственной формулировке, как максимизируемая комбинация внутрикластерных близостей с дисперсией (квадратами расстояний) между кластерами. Показано, что качество кластеризации удастся улучшить по сравнению с обычным критерием качества кластеризации [1].

Работа поддержана грантом РФФИ № 13-07-00010.

- [1] *Двоенко С. Д.* Двухкомпонентная функция качества кластеризации множества элементов, представленных парными сравнениями // «Машинное обучение и анализ данных», 2014 (в печати).

Bipartial objective function for clustering of a set of elements in terms of pairwise comparisons

Dvoenko Sergey

dsd@tsu.tula.ru

Tula, Tula State University

In a featureless case, a set of objects is represented only by results of pairwise mutual comparisons in the form of a distance or similarity matrix.

New meaningless versions of the well-known k -means algorithm are proposed where it is not necessary to calculate cluster centers at all. The k -meanless versions of the k -means algorithm make permutations on the comparison matrix and give the same result for both featureless and feature-based cases.

New bipartial objective function combines intracluster distances with intercluster similarities and needs to be minimized or in the dual form combines intracluster similarities with intercluster distances and needs to be maximized. It is shown that the clustering quality can be improved compared with the usual objective function [1].

This research is funded by RFBR, grant 13-07-00010.

- [1] Dvoenko S. 2014 (in press). Bipartial objective function for clustering of a set of elements in terms of pairwise comparisons. *Machine Learning and Data Analysis*.

Оптимальная коррекция метрических нарушений в матрицах парных сравнений

Двоенко Сергей Данилович

sergedv@yandex.ru

*Пшеничный Денис Олегович**

denispshenichny@yandex.ru

Тула, Тульский государственный университет

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

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

В работе предлагается способ коррекции нормированных матриц парных близостей с тем, чтобы скорректированная матрица была положительно определенной и минимально отличающейся от исходной.

Предлагаемый метод основан на выявлении объектов, вносящих нарушения в метрику, и последующей коррекции их парных сравнений с другими элементами множества, причем данный подход позволяет выбирать набор этих элементов. Показано, что такая коррекция всегда существует и может быть оптимальной.

В отличие от традиционного подхода, основанного на дискретном разложении Карунена–Лоэва, предложенный метод позволяет корректировать не все, а лишь некоторые элементы матрицы парных сравнений [1].

Работа поддержана грантом РФФИ № 13-07-00010.

- [1] Двоенко С. Д., Пшеничный Д. О. Оптимальная коррекция метрических нарушений в матрицах парных сравнений // Машинное обучение и анализ данных, 2014. Т. 1. № 7. С. 885–890. <http://jmla.org/papers/doc/2014/no7/Dvoenko2014Optimal.pdf>.

Optimal correction of metrical violations in matrices of pairwise comparisons

Dvoenko Sergey

sergedv@yandex.ru

*Pshenichny Denis**

denispshenichny@yandex.ru

Tula, Tula State University

In modern data mining, experimental data are usually represented as objects' mutual pairwise comparisons.

The condition of correct immersion of set objects in metrical space in the absence of the initial feature space is nonnegative definiteness of matrix of pairwise comparisons between these objects. In this case, similarities are interpreted as scalar products and dissimilarities as distances.

This paper suggests a method of correction for normalized matrices of similarities; so, the corrected matrix is positively definite and minimally deviated from the initial one.

The proposed method detects objects, which contribute violations in the metrics. Pairwise comparisons of these objects with subset of other objects are corrected. This approach also allows to choose this subset of elements. It is proved that such a correction always exists and can be optimal.

In contrast to traditional approach based on Karhunen–Loeve discrete decomposition, the proposed method is able to correct only few of matrix elements [1].

This research is funded by RFBR, grant 13-07-00010.

- [1] Dvoenko S., Pshenichny D. 2014. Optimal correction of metrical violations in matrices of pairwise comparisons/ *Machine Learning and Data Analysis*. 1(7):885–890. <http://jmla.org/papers/doc/2014/no7/Dvoenko2014Optimal.pdf>.

Выбор метрики в задачах частичного обучения

Иофина Галина Владимировна¹

giofina@mail.ru

Минаев Андрей Викторович¹

andrej.minaev@phystech.edu

Поляков Юрий Сергеевич¹

polyakovyury@gmail.com

Максимов Юрий Владимирович^{1,2*} yury.maximov@phystech.edu

¹Москва, Московский физико-технический институт (МФТИ)

²Москва, Институт проблем передачи информации РАН

Рассматриваются задачи машинного обучения с частичной информацией (semisupervised learning). Предполагается, что входом задачи является простая выборка $(x_i, y_i)_{i=1}^u$ из распределения D на множестве $\mathcal{X} \times \mathcal{Y}$, $|\mathcal{Y}| \leq \infty$. При этом значения y_{n+1}, \dots, y_u алгоритму классификации не известны.

Задача частичного обучения состоит в том, чтобы гарантировать восстановление классификации на неизвестных алгоритму объектов из того же распределения с избыточным риском ε и достоверностью δ при наличии минимального числа размеченных $n(\varepsilon, \delta)$ и неразмеченных $u(\varepsilon, \delta)$ объектов.

В работе предлагается усиление последних результатов Unger *et al.* об оценке числа размеченных объектов $n(\varepsilon, \delta)$, достаточных для решения поставленной задачи. Результаты Unger *et al.* базируются на применении двухэтапной схемы классификации, при этом на первом этапе используется классификация методом ближайших соседей, на втором строится классификатор из заданного семейства. В настоящем докладе, в частности, показано, что правильный выбор метрики в схеме ближайших соседей [1] позволяет существенно усилить наилучшие известные авторам оценки этого типа на $n(\varepsilon, \delta)$.

Исследование выполнено при финансовой поддержке РФФИ в рамках научных проектов № 14-01-00722 а, № 14-07-31241 мол_а и № 14-07-31277 мол_а, а также Лаборатории структурных методов анализа данных в предсказательном моделировании ФУПМ МФТИ, грант правительства РФ дог. 11.G34.31.0073.

- [1] Iofina G. V. A study of metrics in finite sets for application in classification and recognition problems // Pattern Recognition and Image Analysis, 2010. Iss. 4. P. 438–446. <http://link.springer.com/article/10.1134/S1054661810040036>.

A Study of Metrics for Semisupervised Learning Problems

*Iofina Galina*¹

giofina@mail.ru

*Minaev Andrey*¹

andrej.minaev@phystech.edu

*Polyakov Yury*¹

polyakovyury@gmail.com

Maximov Yury^{1,2*}

yury.maximov@phystech.edu

¹Moscow, Moscow Institute of Physics and Technology (MIPT)

²Moscow, Institute for Information Transmission Problems RAS

The reported study is dealing with semisupervised learning problems. It is supposed that the problem's input is a sample $(x_i, y_i)_{i=1}^u$ drawn i.i.d. (independent and identically distributed) from a fixed but unknown distribution D over $\mathcal{X} \times \mathcal{Y}$, $|\mathcal{Y}| \leq \infty$. Assume that algorithm does not have an access to the values y_{n+1}, \dots, y_u .

The problem is to guarantee the excess risk bounds at level ε with confidence δ using as least as possible number of labeled $n(\varepsilon, \delta)$ and unlabeled $u(\varepsilon, \delta)$ objects.

In the report, a slight modification of the two-stage scheme proposed in recent works by R. Urner *et al.*, where on the first stage the nearest neighbor classification was used to increase the size of labeled data, is suggested. Namely, it is proofed that using some ideas from [1], it is possible to improve the quality of this step and decrease the estimates on the number of the needed labeled examples $n(\varepsilon, \delta)$.

The reported study was partially supported by RFBR, research projects 14-01-00722 a, 14-07-31241 mol.a, and 14-07-31277 mol.a and by Laboratory for Structural Methods of Data Analysis in Predictive Modeling, MIPT, RF government grant, ag. 11.G34.31.0073.

[1] Iofina G.V. 2010. A study of metrics in finite sets for application in classification and recognition problems. *Pattern Recognition and Image Analysis* 4:438–446. <http://link.springer.com/article/10.1134/S1054661810040036>.

О комбинаторных оценках максимальных ε -разбиений метрических конфигураций

Пушняков Алексей Сергеевич aleksey.pushnyakov@phystech.edu
Долгопрудный, Московский физико-технический институт

Пусть дано метрическое пространство (X, ρ) с конечным числом точек. Под ε -кластером понимается множество диаметра не более ε . Если в метрическом пространстве есть *достаточно большой* ε -кластер, то расстояний, больших ε , *мало*. Рассматривается обратная задача: пусть известно, что расстояний, больших ε , *мало* в следующем смысле

$$|\{(x, y) : \rho(x, y) > \varepsilon\}| \leq \frac{\delta |X|^2}{2}. \quad (1)$$

Требуется определить, для каких ε' можно гарантировать существование ε' -кластера мощности $(1 + o(1))|X|$ при $\delta \rightarrow 0$.

Оказывается, что в общем случае при $\varepsilon' < 2\varepsilon$ существуют метрики, для которых выполнено (1), но максимальная мощность ε' -кластера не превосходит $\alpha|X|$ для любого наперед заданного $\alpha > 0$. В случае $\varepsilon' \geq 2\varepsilon$ получена наилучшая нижняя оценка на мощность максимального ε' -кластера, асимптотически равная $(1 - \delta/2 + o(1))|X|$.

В некоторых частных случаях значение ε' может быть уменьшено. Так, в случае евклидовой метрики для $\varepsilon' \geq \sqrt{2}\varepsilon$ получена оценка на мощность максимального кластера вида $(1 - 2\sqrt{\delta})|X|$ [1].

- [1] *Пушняков А. С.* О комбинаторных оценках максимальных ε -разбиений метрических конфигураций // Машинное обучение и анализ данных, 2014. Т. 1. № 7. С. 123–132. <http://jmla.org/papers/doc/2014/no7/Pushnyakov2014MetricPartition.pdf>.

On combinatorial bounds for maximal ε -partitions of a finite metric space

Pushnyakov Alexey

aleksey.pushnyakov@phystech.edu

Dolgoprudny, Moscow Institute of Physics and Technology

A finite metric space (X, ρ) is studied. By ε -cluster, a subset of X with diameter at most ε is meant. If a *large* ε -cluster exists, there are *few* distances more than ε . Let the number of distances more than ε satisfy the following inequality:

$$|\{(x, y): \rho(x, y) > \varepsilon\}| \leq \frac{\delta |X|^2}{2}. \quad (1)$$

It is necessary to determine a new ε' such that there always is an ε' -cluster with cardinality $(1 + o(1))|X|$, $\delta \rightarrow 0$.

In case where $\varepsilon' < 2\varepsilon$ for all $\alpha > 0$, there is a finite metric space satisfying (1) that does not contain ε' -cluster with cardinality more than $\alpha|X|$. In case where $\varepsilon' \geq 2\varepsilon$, the best possible lower bound for ε' -cluster cardinality that asymptotically equals $(1 - \delta/2 + o(1))|X|$ is obtained.

In some particular cases, the value of ε' can be decreased. For instance, in case of Euclidean metric for $\varepsilon' \geq \sqrt{2}\varepsilon$, obtained lower bound equals $(1 - 2\sqrt{\delta})|X|$ [1].

- [1] Pushnyakov A. 2014. On combinatorial bounds for maximal ε -partitions of a finite metric space. *Machine learning and data analysis* 1(7):123–132. <http://jmla.org/papers/doc/2014/no7/Pushnyakov2014MetricPartition.pdf>.

О реляционных факторизациях полуметрического конуса

Рудаков Константин Владимирович^{1,2} rudakov@ccas.ru

¹Москва, МФТИ

²Москва, ВЦ РАН

Теоретические построения основываются на использовании, прежде всего, представления о конечных метриках как о точках полуметрического конуса. При этом оказывается существенной оценка «близости» до центральной оси этого конуса (неинформативные пространства изолированных точек). Описания метрик линейными порядками на парах объектов дополняются порядками на тройках объектов, что позволяет существенно более полно использовать исходную реальную информацию. Отметим, что использование более сложных отношений (непорядковых бинарных или отношений более высокой арности) рассматривается как одно из возможных направлений развития подхода.

В роли функционалов оценки «качества» конечных метрических конфигураций используются параметры, связанные с «нарушениями аксиомы треугольника» в том смысле, что расстояние между двумя объектами оказывается строго больше, чем сумма расстояний от некоторого (другого) объекта до двух каких-либо объектов. Важно, что именно эта ситуация не встречается в окрестности центральной неинформативной оси полуметрического конуса. В [1] приводится описание «максимально информативных» с данной точки зрения конфигураций и рассматриваются соответствующие методы коррекции метрик.

Рассматривается также комплекс моделей алгоритмов распознавания, основанных на вычислении оценок, в которых используются порядковые факторизации исходных метрик. Для этого случая получены конструктивные критерии разрешимости и регулярности задач и условия полноты моделей алгоритмов Деза, Журавлёва и Рудакова.

- [1] *Рудаков К. В.* О применении универсальных ограничений при исследовании алгоритмов классификации // Кибернетика, 1988. № 1. С. 1–5.

On relational factorizations of semimetric cone

Rudakov Konstantin^{1,2}

rudakov@ccas.ru

¹Moscow, MIPT

²Moscow, CCAS

Theoretical development relies primarily on representation of finite metrics as points of semimetric cone. In this case, estimating of the “closeness” to the central axis of the cone (noninformative spaces of isolated points) becomes significant. Metric descriptions with linear orders on the pairs of objects is supplemented with orders on the triples of objects. It allows to use initial real information in a considerably more complete manner. Note that using more complicated relationships such as nonorder binary relationships or relationships of higher arity is considered as possible area of the approach development.

The parameters used to assess the “quality” of finite metric configurations are related to the “violation of triangle axiom.” It means that the distance between two objects is strictly greater than the sum of the distances from some (other) object to any two objects. It is important that it is not hold in the neighborhood of central noninformative axis of semimetric cone. The paper [1] describes configurations with “maximum information” from this point of view and considers the corresponding methods of metric correction.

Then, a system of models of recognition algorithms is considered. The recognition algorithms are based on estimates using order factorization of initial metrics. For this case, a constructive criteria for the problem solvability and regularity and completeness conditions of models of the algorithms of Deza, Zhuravlev, and Rudakov are obtained.

[1] Rudakov K. 1988. On application of generic constraints for classification algorithms research. *Kibernetika* 1:1–5.

О критериях ветвления, используемых при синтезе решающих деревьев

Генрихов Игорь Евгеньевич

`ingvar1485@rambler.ru`

Москва, «Рейс ИТ»

Деревья решений являются известным инструментом, используемым при решении задачи распознавания. Процедура построения классического дерева решений представляет собой итерационный процесс. Как правило, для построения внутренней вершины дерева на каждом шаге выбирается признак, наилучшим образом удовлетворяющий некоторому критерию ветвления. По значениям этого признака и осуществляется ветвление, далее указанная процедура повторяется для каждой из ветвей. В зависимости от выбранного признака деревья могут существенно отличаться как по структуре, так и по распознающим качествам.

В данной работе предложен новый критерий ветвления — критерий максимизации доли объектов различных классов (Maximum Differences of Classes, MDC). На модельных данных проанализированы особенности критерия MDC в сравнении с такими известными критериями, как: Gain, GainRatio, Gini Index, Twoing и критерий равномерного разбиения. На большом числе прикладных задач проведено исследование структурных и распознающих свойств решающего дерева в зависимости от применяемого критерия ветвления: глубина дерева, средняя глубина листьев дерева, «сбалансированность» дерева (разница между глубиной и средней глубиной листьев дерева), взвешенная глубина распределения описаний обучающих объектов по листьям дерева, «оптимальность» распределения обучающих объектов по листьям дерева (абсолютная разница между средней глубиной листьев дерева и взвешенной глубиной распределения описаний обучающих объектов по листьям дерева), качество дерева (с помощью метода “leave-one-out” и анализа распределения отступов обучающих объектов), число листьев дерева.

Работа поддержана грантом РФФИ № 13-01-00787.

- [1] *Генрихов И. Е.* О критериях ветвления, используемых при синтезе решающих деревьев // Машинное обучение и анализ данных, 2014. Т. 1. № 8. С. 988–1017. <http://jmla.org/papers/doc/2014/no8/Genrikhov2014Criteria.pdf>.

About splitting criteria used for synthesis of decision trees

Genrikhov Igor

ingvar1485@rambler.ru

Moscow, Race IT

Decision trees are a well known tool used for solution of a recognition problem. The procedure for constructing a classical decision tree is an iterative process. As a rule, to construct an internal tree vertex at every step, a feature best satisfying a certain splitting criterion is chosen. The values of this feature are used to construct the branches of the tree and the procedure is then repeated for each branch. Depending on the feature chosen, the resulting trees can differ widely in the structure and in recognizing qualities.

In this paper, a new of splitting criterion is proposed — criterion of maximization share objects of different classes (Maximum Differences of Classes, MDC). On the model data, particular qualities of criterion MDC are analyzed in comparison with such famous criteria as Gain, GainRatio, Gini Index, Twoing, and criteria of uniform partition. On a large number of real-world tasks, the structural and recognizable properties of a decision tree are investigated depending on the criteria splitting: depth of tree, average depth of tree leaves, “balance” of tree (differences between depth and average depth of tree leaves), weighted depth of the descriptions distribution of training objects on the leaves of tree, “optimal” distribution of training objects on tree leaves (absolute difference between an average depth of tree leaves and a weighted depth of the descriptions distribution of training objects on the leaves of tree), quality of tree (with method “leave-one-out” and analysis margins distribution of training objects), and number of tree leaves [1].

This research is funded by RFBR, grant 13-01-00787.

- [1] Genrikhov I. 2014. About splitting criteria used for synthesis of decision trees. *Machine Learning and Data Analysis*. 1(8):988–1017. <http://jmla.org/papers/doc/2014/no8/Genrikhov2014Criteria.pdf>.

Эффективный параллельный алгоритм дуализации

*Дюкова Елена Всеволодовна*¹

edjukova@mail.ru

Никифоров Андрей Геннадьевич^{2*}

ankifor@gmail.com

*Прокофьев Пётр Александрович*¹

p_prok@mail.ru

¹Москва, ВЦ РАН

²Москва, ВМиК МГУ

Одной из центральных задач дискретной математики является дуализация — построение всех неприводимых покрытий булевой матрицы. Набор столбцов булевой матрицы L называется покрытием, если в подматрице матрицы L , образованной этим набором, нет строки, состоящей из одних нулей. Покрытие называется неприводимым, если никакое его собственное подмножество не является покрытием. Как правило, число неприводимых покрытий растёт экспоненциально с ростом размеров матрицы, поэтому актуальным является использование параллельных вычислений.

Хорошую практическую применимость показывают асимптотически оптимальные алгоритмы дуализации, которые имеют теоретическое обоснование эффективности в типичном случае (on average). Существуют простые и достаточно очевидные схемы распараллеливания этих алгоритмов, главным недостатком которых является неравномерная загрузка процессоров. В [1] предлагается принципиально новая схема распараллеливания асимптотически оптимальных алгоритмов дуализации (\mathfrak{B} -схема).

Пусть $P_j(L)$ — множество неприводимых покрытий матрицы L , для которых j — наименьший номер столбца. \mathfrak{B} -схема основана на статистическом анализе множеств $P_j(L)$. Показывается, что \mathfrak{B} -схема осуществляет равномерную загрузку процессоров и обеспечивает ускорение, близкое к максимально возможному.

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

- [1] Дюкова Е. В., Никифоров А. Г., Прокофьев П. А. Статистически эффективная схема распараллеливания алгоритмов дуализации // Машинное обучение и анализ данных, 2014. Т. 1. № 7. С. 843–853. <http://jmla.org/papers/doc/2014/no7/Nikiforov2014Statistically.pdf>.

Efficient parallel dualization algorithm

Djukova Elena

edjukova@mail.ru

*Nikiforov Andrey*²★

ankifor@gmail.com

Prokofjev Petr

p-prok@mail.ru

¹Moscow, CC RAS

²Moscow, CMC MSU

Dualization is a fundamental problem in discrete mathematics. It is equivalent to irreducible coverings enumeration of a boolean matrix. A set of columns of boolean matrix L is called a covering if each row of L has at least one unit element in these columns. A covering is called irreducible if it includes no other coverings. As a rule, the number of irreducible coverings is exponential in the input matrix size. Therefore, it is essential to use parallel computing.

Asymptotically optimal dualization algorithms have shown a high level of feasibility. The efficiency on average of these algorithms is proven theoretically. There exist simple and obvious parallel schemes for these algorithms. However, such schemes do not provide a balanced load of processing units. A new efficient scheme, called \mathfrak{B} -scheme, is suggested in [1].

Let $P_j(L)$ be a set of irreducible coverings of matrix L , for which j is the smallest number of column. \mathfrak{B} -scheme is based on statistical analysis of sets $P_j(L)$. This scheme is shown to load processors in a balanced way and to provide a close to maximum speedup.

This research is funded by RFBR, grants 13-01-00787-a and 14-07-00819-a.

- [1] Dyukova E., Nikiforov A., Prokofjev P. 2014. Statistically efficient parallel scheme for dualization algorithms. *Machine Learning and Data Analysis*. 1(7):843–853. <http://jmla.org/papers/doc/2014/no7/Nikiforov2014Statistically.pdf>.

Новые асимптотически оптимальные алгоритмы дуализации

Дюкова Елена Всеволодовна*
Прокофьев Петр Александрович
Москва, ВЦ РАН

edjukova@mail.ru
p_prok@mail.ru

Представлены результаты по усовершенствованию и экспериментальному обоснованию асимптотически оптимального подхода к решению дискретных перечислительных задач. Подход нацелен на построение алгоритмов, эффективных в типичном случае. Одной из центральных перечислительных задач является дуализация — построение всех неприводимых покрытий булевой матрицы. Набор столбцов H булевой матрицы L называется неприводимым покрытием, если выполнены условия: (1) подматрица L^H матрицы L , образованная столбцами набора H , не содержит строку вида $(0, 0, \dots, 0)$; (2) L^H содержит каждую из строк вида $(1, 0, 0, \dots, 0, 0)$, $(0, 1, 0, \dots, 0, 0)$, \dots , $(0, 0, 0, \dots, 0, 1)$. Набор столбцов H , удовлетворяющий условию (2), называется *совместимым*. Асимптотически оптимальные алгоритмы дуализации основаны на перечислении с полиномиальной задержкой «максимальных» совместимых наборов столбцов булевой матрицы. Множество решений пополняется, если для найденного на текущем шаге набора столбцов выполнено условие (1) и на предыдущих шагах этот набор не строился. В противном случае шаг алгоритма считается «лишним». Эффективность алгоритма обосновывается тем, что число его лишних шагов для почти всех булевых матриц данного размера имеет более низкий порядок роста по сравнению с числом всех неприводимых покрытий матрицы с ростом ее размера.

В [1] построены асимптотически оптимальные алгоритмы дуализации RUNC, RUNC-M и PUNC. Показано, что предлагаемые алгоритмы требуют меньших временных затрат по сравнению с асимптотически оптимальными алгоритмами, построенными ранее в отечественных и зарубежных публикациях.

Работа частично поддержана грантами РФФИ № 13-01-00787-а и №14-07-00819-а.

- [1] Дюкова Е. В., Прокофьев П. А. Построение и исследование новых асимптотически оптимальных алгоритмов дуализации // Машинное обучение и анализ данных, 2014. Т. 1. № 8. С. 1048–1067. <http://jmla.org/papers/doc/2014/no8/Prokofjev2014Construction.pdf>.

New asymptotically optimal algorithms for dualization

Djukova Elena *

edjukova@mail.ru

Prokofjev Petr

p-prok@mail.ru

Moskow, Dorodnicyn Computing Centre of RAS

An improvement and experimental justification of the asymptotically optimal approach to discrete enumeration problems are presented in [1]. The approach is aimed at effective on average algorithms construction.

Dualization is considered as a fundamental enumeration problem. It is equivalent to irreducible coverings enumeration of a given boolean matrix. A column set H of a boolean matrix L is called an irreducible covering if (1) submatrix L^H formed by columns from H contains no zero-filled rows; and (2) L^H contains each row from the list: $(1, 0, 0, \dots, 0, 0), (0, 1, 0, \dots, 0, 0), \dots, (0, 0, 0, \dots, 0, 1)$. Column set H satisfying the condition 2) is called *compatible*.

Asymptotically optimal dualization algorithms enumerate with polynomial delay the “maximal” compatible column sets. Such algorithms construct a column set at each step. The solutions set is updated if the column set satisfies condition (1) and has not been found at previous steps. Otherwise, this step is considered “extra”. The proportion of “extra” steps tends to zero for almost all boolean matrices of a given size as matrix size increases.

Asymptotically optimal dualization algorithms RUNC, RUNC-M and PUNC are suggested in [1]. These algorithms are shown to outperform the prior asymptotically optimal ones.

This study was financially supported by the Russian Foundation for Basic Research (projects 13-01-00787-a and 14-07-00819-a).

- [1] Djukova E. V., Prokofjev P. A. 2014. Construction and investigation of new asymptotically optimal algorithms for dualization. *Machine Learning and Data Analysis*. 1(8):1048–1067. <http://jmla.org/papers/doc/2014/no8/Prokofjev2014Construction.pdf>.

Методы интеллектуальной обработки качественных данных

Покровская Ирина Вячеславовна

ivp750@mail.ru

Гольдовская Марина Дмитриевна

mdgold54@mail.ru

Дорофеев Юлиа Александровна

dorofeyuk_julia@mail.ru

Киселёва Нелли Евсеевна

nellkis@mail.ru

Москва, ИПУ РАН

Исследованы задачи интеллектуальной обработки качественных данных. Рассмотрено два примера постановок задач и алгоритмов обработки качественных данных, представленных в виде признаков долевого типа и эмпирических графов большой размерности. Разработана методика интеллектуальной обработки признаков долевого типа, проведено тестирование на реальных данных. Исследованы возможности точного и приближенного представления графа большой размерности через его описание. На задачу агрегирования был распространен оптимизационный подход к построению размытой классификации. В рамках структурно-классификационной методологии интеллектуального анализа сложно организованных данных разработаны оригинальные алгоритмы решения задачи обработки информации с помощью агрегирования графов большой размерности [1].

Работа поддержана грантами РФФИ № 14-07-00463-а, № 13-07-00992-а и № 12-07-00540-а.

- [1] *Покровская И. В., Гольдовская М. Д., Дорофеев Ю. А., Киселёва Н. Е.* Методы интеллектуальной обработки качественных данных // Машинное обучение и анализ данных (в печати).

Intellectual processing methods of qualitative data

*Pokrovskaya Irina**

ivp750@mail.ru

Goldovskaya Marina

mdgold54@mail.ru

Dorofeyuk Yulia

dorofeyuk_julia@mail.ru

Kiseleva Nelli

nellkis@mail.ru

Moscow, ICS RAS

Intellectual processing of qualitative data problem was investigated. Two examples of the states of the problems and algorithms for qualitative data processing, presented in the form of the equity-type characteristics and large-dimension empirical graphs were considered. The methodology of data mining (group) characteristics of equity-type (equivalent blurred classifications) was developed, this method was tested on real data. The possibilities of the exact and approximate representation of the large-dimension graph through its description were studied. The optimization approach to the construction of the fuzzy classification was distributed to the problem of aggregation. In the framework of the structural-classification mining methodology of complex data, the original information processing algorithms by large-dimension graphs aggregation methods were developed.

This research is funded by RFBR, grants 14-07-00463, 13-07-00992, and 12-07-00540.

- [1] Pokrovskaya I. V., Goldovskaya M. D., Dorofeyuk J. A., Kiseleva N. E. 2014 (in press). Intellectual processing methods of qualitative data. *Machine Learning and Data Analysis*.

Инварианты размеченных графов в комбинаторной теории разрешимости

Торшин Иван Юрьевич^{1*}

tiy135@yahoo.com

Рудаков Константин Владимирович^{1,2}

rudakov@ccas.ru

¹Москва, МФТИ

²Москва, ВЦ РАН

Комбинаторная теория разрешимости, представляющая собой развитие алгебраического подхода к задачам распознавания, является современным инструментом для исследования признаков объектов. В докладе будет представлен формализм, позволяющий применять комбинаторную теорию разрешимости к теоретико-графовым построениям. Введены особые виды разметок — χ -цепи (цепи размеченных вершин) и χ -узлы (подграфы окрестностей размеченных вершин). Показано, что отношения вхождения χ -цепей и χ -узлов в размеченный граф являются инвариантами; получены критерии полноты соответствующих инвариантов. В рамках комбинаторной теории разрешимости размеченные графы рассматриваются как объекты, а их инварианты — как признаковые описания объектов. Получены критерии локальной полноты исследуемых наборов инвариантов (т. е. полноты относительно заданного множества прецедентов). Показано, что комбинаторное тестирование критерия регулярности соответствующей задачи распознавания позволяет количественно оценить локальную полноту используемых инвариантов. Представлены результаты практического применения предлагаемого формализма к ряду прикладных задач [1, 2].

Работа частично поддержана грантами РФФИ №№ 12-07-00485, 13-07-12053 и 14-07-00852.

- [1] *Torshin I. Yu., Rudakov K. V.* On the application of the combinatorial theory of solvability to the analysis of chemographs // Pattern Recognition and Image Analysis, 2014. Vol. 24. P. 11–23.
- [2] *Torshin I. Yu., Rudakov K. V.* Local completeness of the chemographs' invariants in view of the combinatorial theory of solvability // Pattern Recognition and Image Analysis, 2014. Vol. 24. P. 196–208.

Invariants of labeled graphs in the combinatorial theory of solvability

*Torshin Ivan*¹★

tiy135@yahoo.com

Rudakov Konstantin^{1,2}

rudakov@ccas.ru

¹Moscow, MIPT

²Moscow, CC RAS

Combinatorial theory of solvability, which is a development of the algebraic approach to the problems of pattern recognition and classification, is a modern tool for the study of the feature descriptions of objects. The report presents a formalism that allows to apply combinatorial theory of solvability to graph-theoretic constructions. Introduced special kinds of markup are χ -chains (chains of labeled vertices) and χ -nodes (subgraphs of neighborhoods of labeled vertices). It is shown that the relation of occurrence of χ -chains and χ -nodes are graph invariants. Completeness criteria of corresponding invariants were obtained. In the framework of the combinatorial theory of solvability, the labeled graphs are treated as objects and their invariants as features of the objects. The criteria for the local completeness of the investigated sets of invariants (i. e., completeness relative to a given set of precedents) were derived. It is shown that combinatorial testing of the regularity criterion of the corresponding recognition problem quantifies local completeness of the invariants used. The results of practical application of formalism to a number of applications are discussed [1, 2].

This work was supported in part by RFBR grants 12-07-00485, 13-07-12053, and 14-07-00852.

- [1] Torshin I. Yu., Rudakov K. V. 2014. On the application of the combinatorial theory of solvability to the analysis of chemographs. *Pattern Recognition and Image Analysis* 24:11–23.
- [2] Torshin I. Yu., Rudakov K. V. 2014. Local completeness of the chemographs' invariants in view of the combinatorial theory of solvability. *Pattern Recognition and Image Analysis* 24:196–208.

Когнитивные средства в интеллектуальных системах поддержки принятия решений

Янковская Анна Ефимовна^{1,2,3*}

ayuankov@gmail.com

*Ямшанов Артем Вячеславович*³

yav@keva.tusur.ru

*Кривдюк Наталья Михайловна*³

skratnat@gmail.com

¹Томск, Томский государственный архитектурно-строительный университет

²Томск, Томский государственный университет

³Томск, Томский государственный университет систем управления и радиоэлектроники

Когнитивные средства широко используются в информационных технологиях для решения различных задач в таких областях, как медицина, образование, геология, проектирование, социология, психология, психиатрия, экобиомедицина, экогеология.

Описан метод матричного представления данных и знаний в интеллектуальных системах поддержки принятия решений (ППР). Под закономерностями понимаются подмножества признаков с определенными легко интерпретируемыми свойствами, влияющими на различимость объектов из разных образов, устойчиво наблюдаемыми для объектов из обучающей выборки и проявляющимися на других объектах той же природы, а также весовые коэффициенты признаков, используемые при ППР.

Представлены оригинальные когнитивные средства: визуализации информационных структур; круговая диаграмма для визуализации закономерностей; 3-симплекс для визуализации ППР и траектории обучения. Используя 3-симплекс, можно представить разбиение динамик тех или иных процессов на подгруппы, что также является преимуществом применения 3-симплекса, особенно для процесса обучения. Описано новое когнитивное средство «Мишень с отображением дополнительных зависимостей» для ППР. Даны области применения когнитивных средств [1].

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

[1] *Yankovskaya A. Yamshanov A. Krivdyuk M.* Application of cognitive graphics tools in intelligent systems // IJEIT, 2014. Vol. 3. Iss. 7. P. 58–65.

Cognitive tools in intelligent systems of decision-making support

Yankovskaya Anna^{1,2,3*}

ayyankov@gmail.com

*Yamshanov Artem*³

yav@keva.tusur.ru

*Krivdyuk Natalia*³

skratnat@gmail.com

¹Tomsk, Tomsk State University of Architecture and Building

²Tomsk, Tomsk State University

³Tomsk, Tomsk State University of Control Systems and Radioelectronics

Cognitive tools are widely used in information technology for solving various problems in such areas as medicine, education, geology, engineering, sociology, psychology, psychiatry, ecobiomedicine, and ecogeology.

The implemented method of matrix representation of data and knowledge in intelligent systems of decision-making support is described. Regularities are subsets of features with particular, easy-to-interpret properties that affect the distinguishability of objects from different patterns that are stably observed for objects from the learning sample and are exhibited in other objects of the same nature and weight coefficients of features used for a decision-making support.

Original cognitive tools: visualization of information structure; circle diagram of regularities visualization; 3-simplex for visualization of decision-making support and the learning trajectory are given. In addition to direct display of process dynamics with using the 3-simplex, it is possible to present partitioning of dynamics of different processes under investigation into subgroups, which is also an advantage of using the 3-simplex and especially important in the process of teaching and learning. New cognitive tool “Target with Display of Additional Dependencies” for decision-making support is proposed. Application areas of cognitive tools are given [1].

The research is supported by the Russian Foundation for Basic Research (projects Nos. 13-07-00373a, 13-07-98037-r_sibir_a, and 14-07-00673) and partially by the Russian Humanitarian Scientific Foundation (project No. 13-06-00709).

[1] Yankovskaya A. Yamshanov A. Krivdyuk M. 2014. Application of cognitive graphics tools in intelligent systems. *IJEIT* 3(7):58–65.

Описание многомерных динамических процессов на языке иерархии концептов

Загоруйко Николай Григорьевич^{1,2} zag@math.nsc.ru
Борисова Ирина Артемовна^{1,2}* biamia@mail.ru
*Леванов Дмитрий Александрович*¹ levanov@gmail.com

¹Новосибирск, Институт математики СО РАН

²Новосибирск, КТИ ВТ СО РАН

Рассмотрение задач анализа данных, возникающих при изучении систем и процессов, описываемых таблицами типа «объект–свойство–время» (кубов данных) привело к заключению о целесообразности описания этих данных с помощью концептов разного иерархического уровня. Концептом здесь называется отрезок времени, на протяжении которого параметры динамического процесса подчиняются заданной закономерности. Переход к описанию некоторой системы в терминах концептов позволяет представить поведение этой системы в упрощенном виде. Для формирования концептов, их типизации и распознавания в потоке данных были разработаны методы, основанные на использовании функции конкурентного сходства (function of rival similarity, FRiS-функции).

Полученная таким образом система описаний поведения объектов применялась для решения задачи распознавания состояний, в которой необходимо сформировать решающее правило, позволяющее различать исследуемые состояния и переходы объектов из одного состояния в другое. Авторами предлагается формальная постановка задачи распознавания состояний объектов и простейшие подходы к ее решению как с использованием предварительной разбивки потока данных на концепты, так и без. Эти подходы тестируются на прикладной задаче из области социальной медицины, посвященной контролю за поведением пациентов по сигналам датчиков, укрепленных на его теле. Результаты эксперимента показывают, что использование FRiS-функции позволяет надежно членить поток данных на участки, соответствующие концептам, и распознавать типы стационарных концептов [1].

Работа поддержана грантом РФФИ № 14-01-00039.

- [1] *Загоруйко Н. Г., Борисова И. А., Леванов Д. А.* Описание многомерных динамических процессов на языке иерархии концептов // Научный вестник НГТУ, 2013. №1(50). С. 60–67.

Description of multivariate dynamic processes with concepts hierarchy

Zagoruiko Nikolay^{1,2}

zag@math.nsc.ru

Borisova Irina^{1,2}★

biamia@mail.ru

Levanov Dmitry¹

levanov@gmail.com

¹Novosibirsk, Institute of Mathematics SB RAS

²Novosibirsk, DTI DT SB RAS

Analyzing the structure of Data Mining tasks, described by data set “object–feature–time” (data cube), gives an idea to describe such data with concepts hierarchy. A concept is interpreted as a time range where dynamic process parameters are conformable to some regularities. After the concepts vocabulary forming, one can transform a dynamic process into text scenario of the process. For concepts extraction, clustering, and recognition, the methods, based on the function of rival similarity, were proposed.

Description of dynamic behavior of a system in terms of concepts hierarchy is used for states recognition problem where different states of objects should be classified. In the work, definition of the problem and simplest methods for its solving are presented. These methods were tested on the task of patient’s activity recognition, where 11 states of the person should be detected on 4 sensors indications. The results of the experiments proved what the methods allow to determine the bounds of stationary concepts and to recognize their types [1].

This research is funded by RFBR, grant 14-01-00039.

- [1] Zagoruiko N., Borisova I., Levanov D. 2013. Description of multivariate dynamic processes with concepts hierarchy. *Naychni Vestnik NGTU* 1(50):60–67.

Использование избыточности в кубах данных для обнаружения ошибок и заполнения пробелов

Загоруйко Николай Григорьевич^{1*} zag@math.nsc.ru
*Татарников Вадим Владимирович*² vadim.tatarnikov@gmail.com

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

²Новосибирск, Новосибирский государственный университет

Рассматриваются возможности использования избыточности информации, содержащейся в кубах данных типа «объект–признак–время», для заполнения пробелов и обнаружения ошибок. Для работы с каждым элементом используется информация не из всего куба данных, а лишь из его «компетентного» подкуба. Назовем строку, столбец и линию, на пересечении которых находится пробел, целевыми фрагментами. В компетентный подкуб входят фрагменты (строка, столбец и линия), наиболее похожие на целевые фрагменты. В качестве меры сходства между фрагментами используется тернарная функция конкурентного сходства (FRiS-функция). Формирование подкуба начинается с создания его «зародыша» в виде пересечения $k \times k \times k$ фрагментов. Затем размерность подкуба увеличивается путем присоединения n_1 наиболее компетентных фрагментов и исключения n_2 наименее компетентных фрагментов, $n_1 > n_2$. Эта процедура Addition–Deletion уменьшает вероятность попадания в локальный экстремум. Процесс наращивания размера подкуба останавливается, когда среди фрагментов, не входящих в подкуб, не находится фрагмента с компетентностью, превышающей заданный порог. Предсказание значения пробела делается с использованием многомерной регрессии между целевыми фрагментами и фрагментами, входящими в состав подкуба. Для оценки ожидаемой ошибки прогноза используется среднее значение ошибок предсказания этим же способом всех известных элементов целевых фрагментов. Приводится пример использования алгоритма для решения прикладной задачи из области нефтедобычи [1].

- [1] *Загоруйко Н. Г., Татарников В. В.* Обнаружение ошибок и заполнение пробелов в кубах данных // Сибирский журнал индустриальной математики, 2014. Т. XVII. № 2(58). С. 50–58.

Use of redundancy in cubes of data for detection of errors and filling of blanks

Zagoruiko Nikolay¹*

zag@math.nsc.ru

Tatarnikov Vadim²

vadim.tatarnikov@gmail.com

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

²Novosibirsk, Novosibirsk State University

The opportunities to use of redundancy of the information containing in cubes of data of type “object–attribute–time” for filling blanks and detection of errors are surveyed. For work with each element, the information not from all cube of data but only from its “competent” subcube is used. Let call a line, a column, and a series, on which crossing there is a blank, target fragments. In competent subcube, the fragments (a line, a column, and a series), which are most similar to target fragments, enter. As a measure of similarity between fragments, figurative function rival similarities (FRiS-function) is used. Formation of subcube starts with creation of its “germ” in the form of crossing $k \times k \times k$ fragments. Then, dimension of the subcube is enlarged by addition of the most competent fragments n_1 and deletion of the least competent fragments n_2 , $n_1 > n_2$. The procedure Addition–Deletion reduces probability of hit in a local extremum. The process of escalating of the subcube stops, when among the fragments which do not enter in subcube, there is no fragment with the competence exceeding the threshold. The prediction of value of a blank is done with use of multivariate regress between target fragments and the fragments which are a part of subcube. For an estimation of an expected error of the forecast, the average value of errors of a prediction by the same way of all known elements of target fragments is used. The ways of acceleration of programs working with big data are described. The example of use of the described program 3D-ZET for the decision of an applied task from area of oil extracting is presented [1].

- [1] Zagoruiko N.G., Tatarnikov V.V. 2014. Discovering of errors and filling gaps in data cubes. *Siberian J. Industrial Mathematics* XVII(2(50–58)).

Синтез линейных многомерных систем квазиинвариантного управления методами распознавания образов

*Теклина Лариса Григорьевна**

teklina2010@yandex.ru

Котельников Игорь Вячеславович

neymark@pmk.unn.ru

Нижегородский государственный университет им. Н.И. Лобачевского

Работа посвящена проблеме синтеза систем управления, которые должны не устранять возникающие ошибки, а предотвращать их, сделав объект управления невосприимчивым (инвариантным) к внешним воздействиям. Рассматривается задача параметрического синтеза линейных многомерных (ММО — Multiple Input Multiple Output) систем квазиинвариантного управления с регулированием по ошибке управления. Под квазиинвариантностью понимается требуемая малость ошибки управления в установившемся режиме. Эта задача не является экстремальной и отличается от таковой как по постановке, так и по методам ее решения. Для синтеза таких систем предлагается использовать новый подход, разработанный для синтеза линейных одномерных (SISO — Single Input Single Output) систем квазиинвариантного управления с заданными свойствами. Этот подход основан на постановке и решении задачи синтеза методами распознавания образов с активным экспериментом. Методы распознавания работают в пространствах большой размерности и позволяют найти нужную область искомых параметров с заданной степенью статистической достоверности. Возможности и особенности применения нового подхода проиллюстрированы на примере двух математических моделей систем управления: управление гироскопом и гашение колебаний высотных сооружений [1].

Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации в рамках базовой части.

- [1] Котельников И. В., Теклина Л. Г. Синтез линейных ММО-систем квазиинвариантного управления методами распознавания образов // Вестник ННГУ им. Н. И. Лобачевского, 2014. Вып. 4.

Synthesis of linear multidimensional systems for the quasi-invariant control using methods of pattern recognition

*Teklina Larisa**

teklina2010@yandex.ru

Kotel'nikov Igor

neymark@pmk.unn.ru

Nizhni Novgorod, Lobachevsky State University of Nizhni Novgorod

The work is devoted to the synthesis of the control systems which should not eliminate errors that occur but provide the perturbation invariance for the control object. The problem of parametric synthesis for linear multidimensional (MIMO — Multiple Input Multiple Output) systems of quasi-invariant control with regulation in control error is examined. Under quasi-invariance, it is understood the required little value of the control error in the steady state. This task is not extremal and differs from the same by both the formulation and the methods of its solution. To solve this problem, it is proposed to use a new approach that should be developed for synthesis of the linear one-dimensional (SISO — Single Input Single Output) quasi-invariant control systems with the given properties. This approach is based on the formulation and solution of the synthesis problem as a problem of pattern recognition with the active experiment. The methods of pattern recognition work in spaces of high dimension and allow one to find the desired region of the unknown parameters with the required degree of statistical reliability. The possibilities and peculiarities of the new approach are illustrated by two mathematical models of the control systems: the gyro-platform control and vibration damping of high-rise buildings [1].

This work was financially supported by the Ministry of Education and Science of the Russian Federation within the framework of the base portion.

- [1] Kotel'nikov I., Teklina L. 2014. Synthesis of linear MIMO–systems for the quasi-invariant control using methods of pattern recognition. *Vestnik of Lobachevsky State University of Nizhni Novgorod* 4.

О некоторых вопросах анализа пучков временных рядов

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

¹Москва, САС Институт

²Москва, НИЯУ МИФИ

В настоящей работе рассматривается разрабатываемый авторами подход к поиску закономерностей в пучках нестационарных k -значных временных рядов. Этот подход позволяет выявлять закономерности, которые подвергаются «плавным» структурным изменениям с течением времени.

Настоящая работа посвящена описанию результатов апробации рассматриваемого подхода на модельных и реальных задачах. Испытания на модельных задачах показали, что подход позволяет эффективно находить заложенные закономерности при достаточно высоком уровне шума. Эксперименты на модельных пучках временных рядов показали, что использование меры сходства закономерностей в функционале качества существенно повышает точность прогнозирования. В рамках экспериментов был получен диапазон весов, при котором достигается максимальное качество распознавания. Анализ реальных временных рядов с применением рассматриваемого алгоритма свидетельствовал об эффективности алгоритма при краткосрочном прогнозировании. Вместе с тем алгоритм решает и задачу интеллектуального анализа данных, предложив закономерности, описывающие взаимосвязь одномерных временных рядов.

Таким образом, апробация рассматриваемого подхода к прогнозированию процессов с плавно меняющимися закономерностями на модельных и реальных данных позволяет судить о достаточной эффективности разрабатываемых авторами алгоритмов при анализе пучков временных рядов с плавно меняющимися закономерностями.

Работа поддержана грантом РФФИ № 13-07-00293.

- [1] *Филипенков Н. В., Петрова М. А.* О некоторых вопросах анализа пучков временных рядов // Машинное обучение и анализ данных, 2014 (в печати).

On the Analysis of Multidimensional Time Series

*Nikolay Filipenkov*¹★

n.filipenkov@mail.ru

*Marina Petrova*²

marina.petrova@mail.ru

¹Moscow, SAS Institute

²Moscow, MEFHI

In this paper an approach for discovering rules in nonstationary finite-valued multidimensional time series is discussed. It allows one to discover rules that are slightly changing their structure over time. A measure of rule similarity is introduced and studied as a weight on the graph of rules.

This paper is focusing on the results of the application of the discussed algorithm to the modeled and real problems. The experiments on the model problems show that the approach allows to mine the hidden rules efficiently even under high noise conditions. The experiments on the modeled multidimensional time series show that using the rules similarity measure in the quality function significantly increases the forecast accuracy. During the experiments the weight range for maximum data mining quality is identified. The analysis of real time series based on the discussed approach show the algorithm's efficiency for short-term forecasting. In addition to that the algorithm is also solving the data mining problem while finding the rules describing the interconnection of the univariate time series.

The application of the discussed approach for forecasting the processes with slightly changing rules on modeled and real data shows the efficiency of the developed algorithms for the analysis of multidimensional time series with slightly changing rules.

This research is funded by RFBR, grant 13-07-00293.

- [1] Filipenkov N. V., Petrova M. A. 2014 (in press). On the Analysis of Multidimensional Time Series. *Machine Learning and Data Analysis*.

Интеллектуальные методы структурного анализа многомерных траекторий

Чернявский Александр Леонидович^{1,*} achern@ipu.ru
Дорофеев Александр Александрович^{1,2} daa2@mail.ru
Дорофеев Юлия Александровна^{1,2} dorofeyuk_julia@mail.ru

¹Москва, ИПУ РАН

²Москва, НИУ ВШЭ

Предложена общая постановка задачи интеллектуального анализа динамических объектов, которая формулируется как задача структурного анализа многомерных траекторий (САМТ) в заданном пространстве исходных параметров. Каждому объекту сопоставляется отрезок дискретной траектории (отрезок многомерного дискретного временного ряда) — набор значений параметров, характеризующий этот объект для некоторой дискретной последовательности моментов времени изучаемого отрезка временного ряда. Изучение структуры множества таких траекторий для всего исходного набора объектов и является целью САМТ. В рамках вариационного подхода к задачам структурно-классификационного анализа разработан алгоритм САМТ, позволяющий выделять классы траекторий, каждый из которых характеризует определенный тип динамики функционирования исследуемых объектов. Проведен теоретический анализ этого алгоритма, в достаточно общих предположениях доказана теорема о его сходимости, обеспечивающая стационарность предельного значения функционала (критерия качества) [1].

Работа поддержана грантами РФФИ № 14-07-00463-а и № 13-07-00992-а.

- [1] *Чернявский А. Л., Дорофеев А. А., Дорофеев Ю. А.* Интеллектуальные методы структурного анализа многомерных траекторий // Машинное обучение и анализ данных, 2014 (в печати).

Intellectual methods of structural analysis of multidimensional trajectories

Chernyavskiy Alexander^{1*}

achern@ipu.ru

Dorofeyuk Alexander^{1,2}

daa2@mail.ru

Dorofeyuk Yulia^{1,2}

dorofeyuk_julia@mail.ru

¹Moscow, ICS RAS

²Moscow, SIU HSE

A general formulation of the dynamic objects intellectual analysis problem, which is formulated as the problem of structural analysis of multidimensional trajectories (SAMT) in a given space of initial parameters. Each object is mapped to a segment discrete trajectory (cut multidimensional discrete time series) that is a set of parameters characterizing the object for some discrete sequence of the time moments of the time series interval under study. The study of the structure of such trajectories set for the entire initial objects set is the SAMT purpose. In the framework of the variational approach to the structural-classification analysis problems, the SAMT algorithm was developed that allows to distinguish the classes of paths, each of which is characterized by a specific dynamics type of the studied objects functioning. A theoretical analysis of this algorithm held and the theorem about its convergence providing stationarity limit value of the functional (quality criteria) was proved in quite general assumptions [1].

This research is funded by RFBR, grants 14-07-00463-a and 13-07-00992-a.

- [1] Chernyavskiy A., Dorofeyuk A., Dorofeyuk Y. 2014 (in press). Intellectual methods of structural analysis of multidimensional trajectories. *Machine Learning and Data Analysis*.

К вопросу о создании самообучающейся системы поддержки принятия решений в области биржевой торговли, основанной на отказоустойчивых подмножествах сигнальных признаков

Янковская Анна Ефимовна^{1,2,3*} auyankov@gmail.com

*Аметов Ринат Винурович*² rin@tsuab.ru

*Китлер Сергей Владимирович*³ svkitler@gmail.com

¹Томск, Томский государственный университет

²Томск, Томский государственный архитектурно-строительный университет

³Томск, Томский государственный университет систем управления и радиоэлектроники

При автоматизации процесса инвестирования основной проблемой являются количество и природа исследуемых данных, имеющих стохастический характер. При этом статистические методы прогнозирования либо неэффективны, либо найденные с их помощью закономерности быстро становятся неактуальными.

Для решения задачи инвестирования предлагается создание самообучающейся интеллектуальной системы (ИС), основанной на логико-комбинаторно-вероятностных методах распознавания образов и отказоустойчивых диагностических тестах. Предлагается ряд оригинальных алгоритмов, повышающих точность распознавания и эффективность алгоритмической торговли в целом, предназначенных для выявления отказоустойчивых подмножеств сигнальных признаков и на их основе поиска временных интервалов тренда, на которых динамика изменения финансовых показателей перестает носить стохастический характер.

Предлагаемая самообучающаяся ИС принятия решений в области биржевой торговли будет обладать несомненными конкурентными преимуществами за счет сочетания различных подходов (конвергенции нескольких наук и научных направлений) [1].

Работа поддержана грантами РФФИ № 13-07-00373 и № 13-07-98037.

[1] *Янковская А. Е., Аметов Р. В.* Логико-комбинаторно-вероятностный подход к анализу фондового рынка // Труды Конгресса по интеллектуальным системам и информационным технологиям «IS-IT'10». Москва: Физматлит, 2010. — С. 344–350.

About creation of self-learning support system for decision-making on stock markets based on fault-tolerant subset of signal features

Yankovskaya Anna^{1,2,3*}

ayyankov@gmail.com

*Ametov Rinat*²

rin@tsuab.ru

*Kitler Sergei*³

svkitler@gmail.com

¹Tomsk, Tomsk State University

²Tomsk, Tomsk State University of Architecture and Building

³Tomsk, Tomsk State University of Control Systems and Radioelectronics

At automation of investment process, the basic issue is a quantity and nature of data under investigation having a stochastic nature. Herewith, a stochastic methods of forecasting are either noneffective or the found regularities quickly become irrelevant.

To resolve the problem of investment, it is suggested to implement self-learning intelligent system based on logical and combinational and probabilistic methods of pattern recognition and fault-tolerant diagnostic tests. Also, a number of original algorithms improving an accuracy of recognition and increasing efficiency of algorithmic trade in general are suggested. The algorithms are designed to reveal fault-tolerant subset of signal features and to find time spans of trend based on them. Dynamics of changes of financial performance on the time spans ceases to have a stochastic nature.

Suggested self-learning intelligent system for decision-making on stock market field will possess undoubted competitive advantages at the expense of combination of various approaches (convergence of several sciences and scientific directions) [1].

This work was supported by RFBR, projects Nos. 13-07-00373 and 13-07-98037.

- [1] Yankovskaya A., Ametov R. 2010. Logical and combinatorial and probabilistic approach to analysis of stock market. *Congress for Intelligent Systems and Information Technologies "IS-IT'10" Proceedings*. Moscow: Fizmatlit. 344–350. [In Russian.]

NP-трудность евклидовой задачи о максимальном разрезе

Агеев Александр Александрович

ageev@math.nsc.ru

Кельманов Александр Васильевич

kelm@math.nsc.ru

*Пяткин Артем Валерьевич**

artem@math.nsc.ru

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

В работе исследуется сложность задачи о разрезе максимального веса в полном взвешенном графе, вершинами которого являются некоторые точки q -мерного евклидова пространства. Рассматриваются два варианта задачи: (1) веса ребер равны евклидовым расстояниям между точками, представляющими вершины (*Euclidean Max-Cut problem*); (2) веса ребер равны квадратам этих расстояний (*Quadratic Euclidean Max-Cut problem*).

Вопрос о статусе сложности этих задач был открытым несколько десятилетий.

Euclidean Max-Cut. Дано множество $\mathcal{X} = \{x_1, \dots, x_N\}$ точек из \mathbb{R}^q . Найти такое разбиение множества \mathcal{X} на два подмножества \mathcal{Y} и \mathcal{Z} , что $\sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} \|y - z\| \rightarrow \max$.

Quadratic Euclidean Max-Cut. Дано множество $\mathcal{X} = \{x_1, \dots, x_N\}$ точек из \mathbb{R}^q . Найти такое разбиение множества \mathcal{X} на два подмножества \mathcal{Y} и \mathcal{Z} , что $\sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} \|y - z\|^2 \rightarrow \max$.

Анализируемые варианты актуальны, в частности, в анализе данных и распознавании образов, в компьютерной геометрии, статистической физике, астрофизике, биологии, проектировании микросхем и телекоммуникационных сетей и др.

Авторами доказано [1], что обе задачи (1) NP-трудны в сильном смысле; (2) не имеют вполне полиномиальной приближенной схемы (FPTAS — fully polynomial time approximation scheme), если $P \neq NP$. Доказательства в обоих случаях схожи и базируются на полиномиальной сводимости известной NP-трудной задачи *Minimum Bisection* для кубических графов.

Работа поддержана грантами РФФИ № 12-01-00090, № 12-01-00093 и № 13-07-00070.

- [1] Агеев А. А., Кельманов А. В., Пяткин А. В. Труднорешаемость задачи о разрезе максимального веса в евклидовом пространстве // Доклады академии наук, 2014. Т. 456. № 5. С. 511–513.

NP-hardness of the Euclidean MAX-CUT problem

Ageev Alexander

ageev@math.nsc.ru

Kel'manov Alexander

kelm@math.nsc.ru

*Pyatkin Artem**

artem@math.nsc.ru

Novosibirsk, Novosibirsk State University, Sobolev Institute
of Mathematics

The problem in which given a complete edge-weighted undirected graph whose vertices are the points of the q -dimensional space is considered. It is required to find a cut of maximum total weight. Two special cases are analyzed where the edge weights are equal (*i*) to the Euclidean distances between points representing the vertices (*Euclidean Max-Cut* problem); and (*ii*) to the squares of these distances (*Quadratic Euclidean Max-Cut* problem).

The complexity status of these problem was unknown for decades.

Euclidean Max-Cut. *Given a set $\mathcal{X} = \{x_1, \dots, x_N\}$ of points in \mathbb{R}^q . Find: a partition of the set \mathcal{X} into two subsets \mathcal{Y} and \mathcal{Z} such that $\sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} \|y - z\| \rightarrow \max$.*

Quadratic Euclidean Max-Cut. *Given a set $\mathcal{X} = \{x_1, \dots, x_N\}$ of points in \mathbb{R}^q . Find: a partition of the set \mathcal{X} into two subsets \mathcal{Y} and \mathcal{Z} such that $\sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} \|y - z\|^2 \rightarrow \max$.*

The considered variants are actual, in particular, in data analysis and pattern recognition, in computer geometry, statistical physics, astrophysics, biology, design of VLSI (very-large-scale integration) and communication networks, etc.

In [1], it was proved that both problems (*i*) are strongly NP-hard; and (*ii*) do not admit fully polynomial time approximation schemes (FPTAS) unless $P=NP$. The proofs are similar and based on polynomial-time reductions from the NP-hard *Minimum Bisection* problem on cubic graphs.

This research is funded by RFBR, grants Nos. 12-01-00090, 12-01-00093, and 13-07-00070.

[1] Ageev A. A., Kel'manov A. V., Pyatkin A. V. 2014. NP-hardness of the Euclidean Max Cut problem. *Dokl. Math.* 456(5): 511–513.

Оптимальная оценка базовой характеристики нестабильности частоты осцилляторов

Борисов Борис Дмитриевич

borisov@laser.nsc.ru

Новосибирск, Институт лазерной физики СО РАН

Относительная нестабильность частоты современных оптических осцилляторов — кандидатов на формирование общей шкалы «частота–время» — достигла значений 10^{-18} . Это качество оценивается базовой характеристикой нестабильности во временной области — специальной дисперсией Аллана (ДА). Дисперсия Аллана широко используется и для классификации основных типов стохастических частотных и фазовых возмущений фликкерного типа с неинтегрируемыми спектральными плотностями мощности. С ростом качества стандартов частоты растут и требования к несмещенности, эффективности и состоятельности оценки самой характеристики нестабильности. В докладе определена истинная (невыборочная) дисперсия оценки ДА, анализируются переменные, влияющие на точность оценки ДА и выявляется фактор, вносящий основной вклад в дисперсию оценки базовой характеристики. Этим фактором является дисперсия оценки математического ожидания частоты на фиксированном интервале времени — единственной величины, входящей в определение самой ДА. Рассматривая систему измерения ДА как систему с конечной памятью и используя расширение теории Винера, полученное Заде и Рагаццини, в докладе приводятся оптимальные алгоритмы сглаживания частотных фликкер-шумов различных типов с учетом краевых эффектов на концах интервала измерений. Эти результаты сравниваются по точности (дисперсии) измерений с измерениями, получаемыми традиционной техникой — с помощью электронно-счетного частотомера. В докладе предлагается схема прецизионной автоматической обработки данных измерений ДА на основе фильтрации Калмана для прецизионного определения наклона ДА и определения типов стохастических возмущений и их источников как альтернатива другим методам классификации [1].

- [1] *Borisov B. D.* Efficient estimation of generator instability characteristics in the time domain // *Optoelectronics, Instrumentation and Data Processing*, 2009. Vol. 45. Iss. 3. P. 282–284.

Efficient estimation of main frequency instability characteristic of oscillators

Borisov Boris

borisov@laser.nsc.ru

Novosibirsk, Institute of Laser Physics SB RAS

The fractional frequency instability of modern optical standards — a candidate to forming a total scale “frequency–time” is down to 10^{-18} . This quality of a frequency standard of any bands is commonly estimated by the main instability characteristics in the time domain — special Allan variance (AV). The AV characteristics is also applied for classification of flicker-types of frequency and phase fluctuations with a nonintegrated spectral density. Achievements in improving modern frequency standards, particularly, optical ones, require more statistical accuracy in estimating AV itself and to identify other factors which may influence the efficiency of the AV estimation. The main factor of an inexactitude is a quality of averaging of frequency fluctuations or the variance of the mean value of the fractional frequency on a limited time interval. Because AV is typically a finite-memory system, for improving statistical accuracy of its estimation, an extension of Wiener’s theory for finite-memory systems by Zadex and Ragazzini is used. The general shape of windowing function of a finite duration for optimal estimation of the mean value of frequency fluctuations against different types of flicker noises was obtained. The proposed estimations are compared to the conventional ones taken with a frequency counter using the same averaging time for flicker-type fluctuations. In report, automatic data processing on the basis of Kalman filtration for precise definition of the AV slope is proposed. The AV could be represented as a structural function of the first order and as a moment of the second order of density function of differences of mean values of frequency fluctuations. The dependence of AV and these alternative frequency instability indexes are ascertained in the time domain. Peculiarities of applying these frequency instability indexes, their merits, and demerits were examined [1].

- [1] Borisov B. D. 2009. Efficient estimation of generator instability characteristics in the time domain. *Optoelectronics, Instrumentation and Data Processing*. 45(3):282–284.

Математическое моделирование универсальной характеристики поворотной-лопастной гидротурбины

Волков Юрий Степанович^{1*}

volkov@math.nsc.ru

*Мирошниченко Валерий Леонидович*¹

miroshn@math.nsc.ru

*Салиенко Александр Евгеньевич*²

sa_cae@yahoo.com

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

²Сызрань, ОАО «Тяжмаш»

Рассматривается задача о построении универсальной характеристики рабочего колеса поворотной-лопастной гидротурбины по результатам энергетических испытаний модельной турбины. Универсальная характеристика является основным документом для выбора параметров натурной гидравлической турбины (диаметр рабочего колеса, частота вращения и др.), чтобы обеспечить наиболее эффективную работу турбины при всех режимах ее эксплуатации на конкретной гидроэлектростанции. Дается описание математического аппарата, примененного для создания математической модели универсальной характеристики рабочего колеса поворотной-лопастной гидротурбины по результатам стендовых энергетических испытаний модельной турбины. В основе предложенного подхода лежат методы аппроксимации многомерных функций по хаотически разбросанным данным, созданные авторами путем модификации и обобщения D^m -сплайнов и мультиквадриков Харди. Приводится пример моделирования по реальным данным на основе созданного комплекса программ.

Работа поддержана программой фундаментальных исследований совместных интеграционных проектов СО РАН и УрО РАН, проект № 32.

- [1] *Волков Ю. С., Мирошниченко В. Л., Салиенко А. Е.* Математическое моделирование универсальной характеристики поворотной-лопастной гидротурбины // Машинное обучение и анализ данных, 2014 (в печати).

Mathematical modeling of hill diagram for Kaplan turbine

Volkov Yuriy^{1*}

volkov@math.nsc.ru

*Miroshnichenko Valery*¹

miroshn@math.nsc.ru

*Salienko Alexander*²

sa.cae@yahoo.com

¹Novosibirsk, Sobolev Institute of Mathematics SB RAS

²Syzran, JSC Tyazhmash

The problem of constructing of a hill diagram for the Kaplan turbine wheel on the power test results of the model turbine was considered. The hill diagram is the basic document for choice of full-scale hydraulic turbine parameters (turbine wheel diameter, rotating frequency, etc.) to ensure the most efficient operation of the turbine at all modes of its operation in a particular hydropower station. A description of the mathematical formalism applied for mathematical modeling of the hill diagram of the Kaplan turbine on the power test results of the model turbine. The basis of the proposed approach are the approximation methods for multidimensional functions at scattered data, The methods are based on modifications and generalizations of D^m -splines and Hardy's multiquadrics. An example of modeling for real data on the basis of the program complex is given [1].

This research is funded by the Ural Branch and the Siberian Branch of RAS, project 32.

- [1] Volkov Yu. S., Miroshnichenko V. L., Salienko A. E. 2014 (in press). Mathematical modeling of hill diagram for Kaplan turbine. *Machine Learning and Data Analysis*.

Стохастическая онлайн оптимизация в случае тяжелых хвостов стохастических градиентов

*Гасников Александр Владимирович*¹ avgasnikov@gmail.com
Крымова Екатерина Александровна^{2*} krymova@phystech.edu

¹Москва, Московский физико-технический институт (ГУ)

²Москва, Институт проблем передачи информации

Рассмотрим задачу стохастической онлайн оптимизации

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

при следующих условиях: $E_{\xi^k} [f_k(x; \xi^k)]$ — выпуклые по x функции; существует такой вектор $\nabla_x f_k(x; \xi^k)$, что

$$E_{\xi^k} (\nabla_x f_k(x; \xi^k) - \nabla_x E_{\xi^k} [f_k(x; \xi^k)]) | \Xi^{k-1} \equiv 0,$$

где Ξ^{k-1} — σ -алгебра, порожденная случайными величинами ξ^1, \dots, ξ^{k-1} ; условие тяжелых хвостов градиентов:

$$P_{\xi^k} [M^{-2} \|\nabla_x f_k(x; \xi^k)\|_{\infty}^2 \geq t | \Xi^{k-1}] = O(t^{-\alpha}), \quad \alpha > 2.$$

Онлайновость постановки задачи допускает, что на каждом шаге k функция f_k может подбираться из рассматриваемого класса функций враждебно по отношению к методу генерации последовательности $\{x^k\}$. Целью работы является распространение оценки метода зеркального спуска на случай тяжелых хвостов: существует такое $C_{\alpha} < 0$, что с вероятностью $1 - \sigma$ [1]

$$\begin{aligned} \sum_{k=1}^N E_{\xi^k} [f_k(x^k; \xi^k)] - \min_{x \in S_n(1)} \sum_{k=1}^N E_{\xi^k} [f_k(x; \xi^k)] &\leq \\ &\leq C_{\alpha} M \left((N \ln n)^{1/2} + \left(\frac{N}{\sigma} \right)^{1/\alpha} \right). \end{aligned}$$

Работа выполнена при поддержке РФФИ (грант 14-01-00722-а); Лаборатории структурных методов анализа данных в предсказательном моделировании МФТИ (грант правительства РФ, договор 11.G34.31.0073); Президента РФ (грант № МК-5285.2013.9).

[1] *Гасников А. В., Нестеров Ю. Е., Спокойный В. Г.* Об эффективности одного метода рандомизации зеркального спуска в задачах онлайн оптимизации // Автоматика и телемеханика. — М.: Наука, 2014. 24 с.

Stochastic online optimization in case of heavy-tailed stochastic gradients

*Gasnikov Aleksandr*¹
Krymova Ekaterina^{2*}

avgasnikov@gmail.com
 krymova@phystech.edu

¹Moscow Institute for Physics and Technology

²Moscow, Institute for Information Transmission Problems

Consider a online stochastic optimization problem

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

under the following conditions: $E_{\xi^k} [f_k(x; \xi^k)]$ are the convex functions in x ; there exists a vector $\nabla_x f_k(x; \xi^k)$ such that

$$E_{\xi^k} (\nabla_x f_k(x; \xi^k) - \nabla_x E_{\xi^k} [f_k(x; \xi^k)]) | \Xi^{k-1} \equiv 0$$

where Ξ^{k-1} is the σ -algebra, generated by random variables ξ^1, \dots, ξ^{k-1} ; heavy tails condition:

$$P_{\xi^k} [M^{-2} \|\nabla_x f_k(x; \xi^k)\|_{\infty}^2 \geq t | \Xi^{k-1}] = O(t^{-\alpha}), \quad \alpha > 2.$$

The online framework implies that on each step k , the function f_k is selected from the set of functions and in the worst case, the method of function selection may hinder the method of sequence $\{x^k\}$ generation. The main goal is to obtain the bounds for mirror descent method in the case of heavy tails: there exists $C_\alpha < 0$ such that with probability $1 - \sigma$ [1]

$$\begin{aligned} \sum_{k=1}^N E_{\xi^k} [f_k(x^k; \xi^k)] - \min_{x \in S_n(1)} \sum_{k=1}^N E_{\xi^k} [f_k(x; \xi^k)] \\ \leq C_\alpha M \left((N \ln n)^{1/2} + \left(\frac{N}{\sigma} \right)^{1/\alpha} \right). \end{aligned}$$

This research is funded by RFBR (grant 14-01-00722-a); Laboratory for Structural Methods of Data Analysis in Predictive Modeling, MIPT (RF Government grant, ag. 11.G34.31.0073); President RF (grant MK-5285.2013.9).

[1] Gasnikov A., Nesterov Y., Spokoiny V.G. 2014. On effectiveness of a randomized mirror descent method for the online optimization problem solution. *Automation and remote control*. Moscow: Nauka. 24 p.

**Об асимптотической разрешимости задачи
 m коммивояжеров на случайных входах,
 неограниченных сверху**

*Гимади Эдуард Хайрутдинович**

gimadi@math.nsc.ru

Истомин Алексей Михайлович

alexeyistomin@gmail.com

Рыков Иван Александрович

rykov@ngs.ru

Цидулко Оксана Юрьевна

tsidulko.ox@gmail.com

Новосибирск, Государственный университет, ИМ СО РАН

Рассматривается следующая задача m -Peripatetic Salesman Problem (m -PSP), которая является естественным обобщением классической задачи коммивояжера: в заданном полном n -вершинном графе $G = (V, E)$ с весовой функцией $w_i : E \rightarrow R^+$, $i = 1, \dots, m$, требуется найти m реберно-несмежных гамильтоновых циклов $H_1, \dots, H_m \subset E$ с минимальным суммарным весом входящих в них ребер.

Предполагается, что ребра гамильтоновых циклов имеют разные весовые функции, причем длины ребер — случайные независимые случайные величины с одинаковой функцией распределения в области $[a_n, \infty]$, $a_n > 0$ [1]. Представлен приближенный алгоритм с временной сложностью $O(mn^2)$, получены оценки относительной точности и вероятности несрабатывания на случайных входах, неограниченных сверху, с функцией распределения, доминирующей экспоненциальное распределение с параметром β_n . В частности, таким распределением является усеченно-нормальное распределение с параметром σ_n .

Проведено обоснование условий асимптотической точности алгоритма в виде $\beta/a_n = O(n^\theta)$ и $m \leq n^{1-\theta}$, где β равно одному из соответствующих параметров β_n или σ_n ; $0 < \theta < 1$.

Работа частично поддержана грантом РФФИ (проект № 12-01-00093) и интеграционным проектом СО РАН и УрО РАН № 7В.

- [1] *Гимади Э. Х., Истомин А. М., Рыков И. А., Цидулко О. Ю.* Вероятностный анализ приближенного алгоритма для решения задачи нескольких коммивояжеров на случайных входных данных, неограниченных сверху. *Труды ИММ УрО РАН*, 2014. Т. 20. № 2. С. 88–98.

On asymptotical solvability for the m -peripatetic salesman problem on random instances unbounded from above

*Gimadi Edward**

gimadi@math.nsc.ru

Istomin Alexey

alexeyistomin@gmail.com

Rykov Ivan

rykov@ngs.ru

Tsidulko Oxana

tsidulko.ox@gmail.com

Novosibirsk, State University, IM SB RAS

The m -Peripatetic Salesman Problem (m -PSP) is considered that is a natural generalization of the classical travelling salesman problem (TSP). Given a complete directed n -vertex graph $G = (V, E)$ with weight functions $w_i : E \rightarrow R^+$, $i = 1, \dots, m$, the problem is to find m edge-disjoint Hamiltonian circuits $H_1, \dots, H_m \subset E$, such that minimize their total weight.

The case of the minimum-weight m -PSP on directed or undirected graph with different weight functions of edges in Hamiltonian circuits is considered when the entries of the distance matrix are some independent identically distributed random variables with the values in the range $[a_n, \infty]$, $a_n > 0$ [1]. The authors construct an approximation algorithm with time complexity $O(mn^2)$ and obtain the performance guarantees for the relative errors and the failure probabilities for random instances with distribution functions that dominate the exponential distribution with a parameter β_n . Truncated normal distribution with parameter σ_n is one of the such distributions.

The conditions of asymptotic optimality of the algorithms are established in the form $\beta/a_n = O(n^\theta)$ and $m \leq n^{1-\theta}$ where β is equal to one of corresponding parameter β_n or σ_n ; $0 < \theta < 1$.

The work partially supported by the RFBR (project No.12-01-00093) and integration project SB RAS and UrB RAS No. 7B.

- [1] Gimadi E. Kh., Istomin A. M., Rykov I. A., Tsidulko O. Yu. 2014. Probabilistic analysis of an approximation algorithm for the m -peripatetic salesman problem on random instances unbounded from above. *Trudy IMM UrO RAN* 20(2):88–98. [In Russian.]

Приближенные алгоритмы поиска в графе несмежных клик минимального суммарного веса

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

*Кельманов Александр Васильевич*¹ kelm@math.nsc.ru

*Пяткин Артем Валерьевич*¹ artem@math.nsc.ru

*Хачай Михаил Юрьевич*² mkhachay@imm.uran.ru

¹Новосибирск, Государственный университет, ИМ СО РАН

²Екатеринбург, Уральский федеральный университет, ИММ Уро РАН

Одна из задач выбора подмножества «похожих» элементов во множестве объектов моделируется следующей экстремальной задачей m -WCP (m -Weighted Clique Problem) [1]. Даны полный неориентированный взвешенный n -вершинный граф G и натуральные числа m, L_1, \dots, L_m такие, что $\sum_{k=1}^m L_k \leq n$. Требуется найти семейство m несмежных клик заданных размеров $\{L_k\}$ с минимальным суммарным весом входящих в них вершин и ребер.

Показано, что задача m -WCP NP-трудна в сильном смысле, как в общем случае так и в частных случаях Metric m -WCP и Quadratic Euclidean m -WCP. Приближенное решение задачи m -WCP строится с использованием точного решения вспомогательной задачи поиска m несмежных звезд с минимальным суммарным весом входящих в них вершин и ребер, получаемое транспортным алгоритмом за время $\mathcal{O}(n^{m+2} \log n)$. Алгоритмы решения задач m -WCP имеют ту же трудоемкость, так что при константном числе искоемых клик их статус полиномиален.

Показано, что задача Metric m -WCP решается с оценкой точности $2\left(1 - \frac{\sum_{k=1}^m S(B_k^*)}{\sum_{k=1}^m L_k S(B_k^*)}\right)$, где $S(B_k^*)$ — суммарный вес вершин и ребер в k -й звезде вспомогательной задачи. Для задачи Quadratic Euclidean m -WCP алгоритм является 2-приближенным. Доказано, что в обеих задачах оценки точности достижимы.

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

[1] *Гимади Э. Х., Кельманов А. В., Пяткин А. В., Хачай М. Ю.* Эффективные алгоритмы с оценками точности для некоторых задач поиска нескольких клик в полном неориентированном графе. Труды ИММ Уро РАН. 2014, Т. 20. № 2. С. 99–112.

Approximation algorithms for finding several disjoint cliques in a complete graph with minimal total weight

Gimadi Edward^{1*}

gimadi@math.nsc.ru

*Kel'manov Alexander*¹

kelm@math.nsc.ru

*Pyatkin Artem*¹

artem@math.nsc.ru

*Khachay Mikhail*²

mkhachay@imm.uran.ru

¹Novosibirsk, State University, IM SB RAS

²Ekaterinburg, Ural Federal University, IMM UrB RAS

One of the problems of finding a subset of similar elements in the set of objects is described by the following extremal problem m -WCP (m -Weighted Clique Problem) [1]. Given a complete undirected weighted n -vertex graph G and natural numbers m, L_1, \dots, L_m such that $\sum_{k=1}^m L_k \leq n$. To find a family of m disjoint cliques of given orders $\{L_k\}$ with minimal total weight of edges and vertices.

It is shown that the m -WCP is NP-hard in strong sense, in the general case and in the special cases of Metric m -WCP and Quadratic Euclidean m -WCP. An approximation algorithm for solving the m -WCP is constructed using the exact solution of auxiliary problem of finding a minimum-weight family of m disjoint stars of corresponding orders L_1, \dots, L_m . This solution can be found by the transportation algorithm in time $\mathcal{O}(n^{m+2} \log n)$. The proposed algorithms for the problems m -WCP have the same time complexity and in the case of constant number of searched cliques, their status is polynomial.

It is proved that for the Metric m -WCP, the approximation guarantee $2 \left(1 - \frac{\sum_{k=1}^m S(B_k^*)}{\sum_{k=1}^m L_k S(B_k^*)} \right)$ where $S(B_i^*)$ is the total weight of the vertices and the edges in the i th star of the auxiliary problem, $i = 1, \dots, m$. For the Quadratic Euclidean m -WCP, the approximation ratio 2 is proved. It is shown that both approximation ratios are attainable.

This research is funded by RFBR (grants 12-01-00090, 12-01-00093, and 13-07-00181) and integration projects SB RAS and UrB RAS (No. 12-01-1017/1 and No. 7B).

- [1] Gimadi E. Kh., Kel'manov A. V., Pyatkin A. V., Khachay M. Y. 2014. Efficient algorithms with performance guarantees for some problems of finding several cliques in complete undirected graph. *Tr. IMM UrO RAN* 20(2):99–112. [In Russian.]

Отображения параллельных алгоритмов на суперкомпьютеры экзафлопсной производительности на основе имитационного моделирования

Глинский Борис Михайлович

gbm@sscc.ru

Марченко Михаил Александрович

marchenko@sscc.ru

Родионов Алексей Сергеевич

alrod@sscc.ru

Караваев Дмитрий Алексеевич

kda@opg.sccc.ru

Подкорытов Дмитрий Игоревич

d.podkorytov@gmail.com

Новосибирск, ФБГУН Институт вычислительной математики
и математической геофизики СО РАН

Проблема исследования свойств масштабируемости параллельных алгоритмов при их реализации на будущих суперЭВМ экзафлопсной производительности выходит за уровень технологических задач. В работе [1] показано, что оценить поведение алгоритмов, разработать модифицированные схемы вычислений можно уже сейчас путем реализации их на имитационной модели. В ИВМиМГ СО РАН разработана система моделирования AGNES, которая была использована для изучения возможности масштабирования распределенного статистического моделирования и решения задачи численного трехмерного моделирования распространения сейсмических волн. Реальными вычислениями показано, что метод Монте–Карло линейно распараллеливается до 1000 вычислительных ядер. Имитационным моделированием исследовано поведение метода до 5×10^5 ядер. Для другой задачи показано хорошее соответствие экспериментальных и модельных результатов до 32 768 ядер. На имитационной модели удалось получить результаты для 1 124 864 ядер. Расчеты проводились на кластерах ЦКП «Сибирский суперкомпьютерный центр».

Работа поддержана грантами РФФИ №№ 13-07-00589 и 14-07-00832, 14-05-00867, МИП 130 СО РАН и Программой РАН 4.9.

- [1] *Глинский Б. М., Марченко М. А., Михайленко Б. Г., Родионов А. С., Черных И. Г., Караваев Д. А., Подкорытов Д. И., Винс Д. В.* Отображения параллельных алгоритмов на суперкомпьютеры экзафлопсной производительности с различными архитектурами на основе имитационного моделирования // Информационные технологии и вычислительные системы, 2013. № 4. С. 3–14.

Mappings of parallel algorithms on supercomputers with ekzaflops performance on the basis of simulation

*Glinsky Boris**

gbm@sscc.ru

Marchenko Mikhail

marchenko@sscc.ru

Rodionov Alexey

alrod@sscc.ru

Karavaev Dmitry

kda@opg.sccc.ru

Podkorytov Dmitry

d.podkorytov@gmail.com

Novosibirsk, Institute of Computational Mathematics and Mathematical Geophysics SB RAS

The problem of investigating the properties of scalability of parallel algorithms for implementing them in future supercomputers with ekzaflops performance goes beyond the level of technological problems. In [1], it is shown that it is possible to estimate the behavior of algorithms and to develop the modified computation scheme by implementing them on a simulation model. In ICMMG, the simulation system AGNES was developed, which was used to study the scalability of distributed statistical modeling and for solving the problem of numerical three-dimensional modeling of seismic wave propagation. Real calculations showed that the Monte-Carlo method is linearly parallelized up to 1000 computing cores. The behavior of the method was investigated with simulation up to 5×10^5 cores. For another problem, it is shown a good compliance between experimental and model results up to 32 768 cores. On the simulation model, the results were obtained for 1 124 864 cores. The calculations were performed on the clusters of CCU “Siberian Supercomputer Center.”

This research is funded by RFBR, grants 13-07-00589, 14-07-00832, and 14-05-00867, IP 130 SB RAS, and Program of RAS 4.9.

- [1] Glinsky B., Marchenko M., Mikhailenko B., Rodionov A., Chernykh I., Podkorytov D., Karavaev D., Weins D. 2013. Simulation modeling of parallel algorithms for different architectures of exaflops supercomputers. *Information Technologies and Computing Systems* 4:3–14.

Вычислительная технология оценки степени выпуклости многоэкстремальной функции

*Горнов Александр Юрьевич**
Зароднюк Татьяна Сергеевна

gornov@icc.ru
tz@icc.ru

Иркутск, Институт динамики систем и теории управления СО РАН

Задачи оптимизации естественно возникают при применении метода математического моделирования. Успех моделирования самым серьезным образом зависит от того, насколько удачно удалось сформировать адекватную модель и, в первую очередь, от выпуклости или невыпуклости привлеченных функций. Можно утверждать, что класс выпуклых функций неплохо освоен математически. Однако ситуация самым существенным образом меняется при рассмотрении невыпуклых задач.

В работе предлагается методика определения степени выпуклости функции, основанная на ее стохастической аппроксимации на всей исследуемой области. Основной идеей подхода является поточечное исследование выпуклости функции по случайно выбранным направлениям и систематизация полученной информации с целью получения интегральной оценки выпуклости. Эффективность предложенной технологии демонстрируется на ряде модельных примеров небольших размерностей, для которых построены и визуализированы области выпуклости функций.

С применением предлагаемой вычислительной технологии возможно производить селекцию используемых при моделировании функциональных зависимостей с целью выбора более удобных для последующего оптимизационного анализа. Работоспособность предложенного подхода исследована на ряде тестовых и модельных задач. Полученные результаты численных экспериментов позволяют надеяться на создание нового вычислительного инструмента, полезного при решении практических задач из различных научно-технических областей [1].

Работа поддержана грантом РФФИ № 12-07-33021 и интеграционным проектом СО РАН № 81.

[1] *Горнов А. Ю., Зароднюк Т. С.* Вычислительная технология оценки степени выпуклости многоэкстремальной функции // Машинное обучение и анализ данных, 2014 (в печати).

Computing technology for estimation of convexity degree of the multiextremal function

*Gornov Alexander**

gornov@icc.ru

Zarodnyuk Tatiana

tz@icc.ru

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

Optimization problems arise in the application of mathematical modeling method. Advance in applying of mathematical modeling depends on how successfully the researcher can construct a valid model and, primarily, on the convexity or nonconvexity of the involved functions. It can be argued that the class of convex functions mathematically well studied. However, the situation greatly changed in case of nonconvex problems.

This paper proposes a technique of determining the degree of the function convexity, based on its stochastic approximation for the considerable area. The main idea of the approach is the pointwise study of the function convexity on the stochastic selected areas and systematization of this information to obtain an integrated estimate of convexity. The effectiveness of the proposed technology is demonstrated on a number of model examples of small dimensions which are constructed and convexity area are visualized.

The selection of functional, used in the mathematical modeling can be produced in order to choose more convenient for optimization analysis with the application of the proposed computing technology. This technique allows to demonstrate “the areas of convexity-nonconvexity” for the problems of small dimensions. The algorithm can be easily parallelized. The efficiency of the considered approach is investigated on a number of test and model problems. The obtained numerical results allow to expect for the creation of a new computational software useful in solving practical problems in various scientific and technical fields [1].

This research is funded by RFBR, grant 12-07-33021, and integration project of SB RAS N 81.

- [1] Gornov A., Zarodnyuk T. 2014 (in press). Computing technology for estimation of convexity degree of the multiextremal function. *Machine Learning and Data Analysis*.

Вычислительная сложность минимизации булевых функций в классе схем глубины 2

*Иофина Галина Владимировна*¹ giofina@mail.ru
Максимов Юрий Владимирович^{1,2*} yury.maximov@phystech.edu

¹Москва, Московский физико-технический институт (МФТИ)

²Москва, Институт проблем передачи информации РАН

Задача минимизации булевых функций в различных классах схем является достаточно хорошо исследованной поли-экстремальной задачей комбинаторной оптимизации. В зависимости от способа задания функции указанная задача может лежать в различных сложностных классах (например, задача минимизации булевой функции в классе схем фиксированной глубины может быть как NP-полной, так и Σ_2^P -полной проблемой).

Учитывая сложностные характеристики проблемы, одним из наиболее популярных направлений исследования в данной области является задача поиска полиномиально разрешимых случаев. К числу таких случаев, в частности, относятся задачи с априорными ограничениями на баланс нулей и единиц (см. работы Ю. И. Журавлева, А. Г. Дьяконова, А. Ю. Когана и др.), либо с ограничением на структуру функции, такими как, например, наличие Horn property (см. работы Е. Voros и др.).

В настоящем докладе будут рассмотрены границы возможностей полиномиальных методов для построения приближенных к минимальным булевым схем реализующих функции с априорными ограничениями, в рамках доклада будет представлен взгляд на проблему минимизации ДНФ специальных функций, изложенную в [1] с точки зрения “классической” теории сложности.

Исследование выполнено при финансовой поддержке РФФИ в рамках научных проектов № 14-07-31241 мол_а и № 14-07-31277 мол_а; а также Лаборатории структурных методов анализа данных в предсказательном моделировании ФУПМ МФТИ, грант правительства РФ дог. 11.G34.31.0073.

- [1] *Maximov Yu. V.* Implementation of Boolean functions with a bounded number of zeros by disjunctive normal forms // *Comput. Math. Math. Phys.*, 2013. Vol. 53. Iss. 9. P. 1391–1409. <http://link.springer.com/article/10.1134/S096554251309008X>.

Computational Complexity of Unbalanced Boolean Minimization within Depth-2 Circuits

*Iofina Galina*¹

giofina@mail.ru

Maximov Yury^{1,2*}

yury.maximov@phystech.edu

¹Moscow, Moscow Institute of Physics and Technology (MIPT)

²Moscow, Institute for Information Transmission Problems RAS

Boolean circuit minimization problem is a well-studied combinatorial optimization problems to attain local minima in a waste amount of points. Its computational complexity status depends on the input. For instance, DNF (disjunctive normal form) minimization could be NP-complete if the function is defined by its truth table or Σ_2^P -complete if the function is defined by a DNF.

Exploration of the various polynomial solvable cases of the problem above is an attractive area of research due to the complexity status of the problem. A number of them were sorted out before (“unbalanced” functions, Horn functions). Nevertheless, a lot of important issues are still uncovered.

Within the report, another point of view on the results, briefly described in [1], will be provided. Namely, lower bounds on approximation rate will be derived for the class of polynomial algorithms for the problem of finding minimal fix-depth circuit of functions attaining the value 0 in a fix (but dimension-dependent) number of points. The present authors’ results rely on exponential time hypothesis, introduced by D. Lokshtanov *et al.*

The reported study was partially supported by RFBR, research projects 14-07-31241 mol_a and 14-07-31277 mol_a and by Laboratory for Structural Methods of Data Analysis in Predictive Modeling, MIPT, RF government grant, ag. 11.G34.31.0073.

- [1] Maximov Yu. V. 2013. Implementation of Boolean functions with a bounded number of zeros by disjunctive normal forms. *Comp. Math. Math. Phys.* 53(9):1391–1409. <http://link.springer.com/article/10.1134/S096554251309008X>.

Быстрые алгоритмы. Представление проекта.

*Карацуба Екатерина Анатольевна*¹

ekar@ccas.ru

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

Целью проекта является развитие теории быстрых вычислений и исследование эффективности внедрения различных быстрых методов и алгоритмов [1] для их дальнейшего результативного использования в программном обеспечении высокопроизводительных вычислительных систем.

Внедрение быстрых алгоритмов в программное и аппаратное обеспечение в новых моделях процессоров, а также в такие пакеты программ, как Maple, Mathematica, Matlab, Arithmos и др., показали, что эти алгоритмы обеспечивают надежные, высокоточные вычисления и потому начинают использоваться компьютерными фирмами в большем объеме. Однако систематическое тестирование быстрых алгоритмов, их сравнение между собой и с обычными методами и алгоритмами, исследования эффективности их внедрения до того не проводились. Быстрые алгоритмы вычисления трансцендентных функций (см. <http://www.ccas.ru/personal/karatsuba/alg.htm>) вообще никогда не тестировались и не сравнивались.

Предполагается параллельно несколько направлений проекта:

1. Описание всех существующих быстрых алгоритмов.
2. Сравнение существующих быстрых алгоритмов для отдельного вычисления путем тестирования основанных на этих алгоритмах программ. По результатам работы создание единого справочника эффективности работы всех быстрых алгоритмов.
3. Усовершенствование программ, основанных на быстрых алгоритмах, для расширения пределов их применимости.
4. Увеличение эффективности работы существующих быстрых алгоритмов, в том числе путем их распараллеливания. Получение новых рекордов в вычислениях.
5. Построение новых быстрых алгоритмов вычислений широкого класса.

- [1] *Карацуба Е. А.* Об одном методе быстрого приближения дзета-констант рациональными дробями // Проблемы передачи информации, 2014. Т. 50. № 2. С. 186–202.

Fast algorithms. Project presentation

*Karatsuba Ekaterina*¹

ekar@ccas.ru

¹Moscow, Dorodnicyn Computing Center of RAS

The aim of the project is the development of the theory of fast computations and the study of the effectiveness of implementation of various fast methods and algorithms [1] for their further effective use in the software of high-performance computing systems.

Implementation of fast algorithms in the software and hardware of new models of processors as well as software packages such as Maple, Mathematica, Matlab, Arithmos, and others demonstrated that these algorithms provide reliable, high-precision calculations and, therefore, modern computer firms begin to use them more often than before. However, systematic testing of fast algorithms and their comparison in the efficiency and also, the comparison of fast algorithms with ordinary methods and algorithms have never been conducted. Fast algorithms for evaluation of transcendental functions (see <http://www.ccas.ru/personal/karatsuba/alg.htm>) have never been tested and compared at all.

The project is assumed to have some parallel directions:

1. A description of all existing fast algorithms.
 2. The comparison of existing fast algorithms for computing a function by testing the corresponding programs. As a result, the creation of a guide in the performance of all fast algorithms.
 3. The improvement of the programs based on fast algorithms to expand the limits of applicability of fast algorithms.
 4. The increasing of efficiency for existing fast algorithms, including via their parallelization. Preparation of new records in the high precision calculations.
 5. The construction of new fast algorithm for computations of wide class functions.
- [1] Karatsuba E. 2014. On one method for fast approximation of zeta constants by rational fractions. *Problems of Information Transmission* 50(2):186–202.

Полиномиальные алгоритмы с гарантированными оценками точности для некоторых задач анализа и распознавания последовательностей

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

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

²Новосибирск, Новосибирский государственный университет

Рассматриваются задачи дискретной оптимизации, которые индуцируются, в частности, проблемами анализа и распознавания последовательностей. Цель доклада — обзор новых и недавних результатов (полученных в Институте математики им. С. Л. Соболева СО РАН) по исследованию вычислительной сложности этих задач и обоснованию полиномиальных алгоритмов с гарантированными оценками точности для решения этих задач. Рассматриваемые задачи имеют приложения в теории приближения, компьютерной геометрии, математической статистике. Они актуальны в широком спектре прикладных проблем. Например, они возникают в медицинской и технической диагностике, биометрике, электронной разведке, космическом мониторинге, дистанционном зондировании Земли и Мирового океана, в обработке экспериментальных данных и т. п. Сущность проблем состоит в поиске информационно значимых подмножеств и подпоследовательностей в зашумленных массивах данных, в разбиении этих данных на подмножества и подпоследовательности, в принятии решения об объектах, порождающих эти данные.

Основные результаты состоят в следующем. Было изучено несколько новых (ранее не изученных), известных (но слабо изученных) и открытых проблем. Кроме того, исследованы некоторые обобщения и специальные случаи классических задач разбиения, а также поиска векторных подмножеств и подпоследовательностей в евклидовом пространстве. Найдены NP-трудные и полиномиально разрешимые случаи этих задач (и соответствующих им проблем анализа данных и распознавания образов). Для их решения предложены точные полиномиальные и псевдополиномиальные алгоритмы, а также полиномиальные алгоритмы с гарантированными оценками точности.

Работа поддержана грантами РФФИ № 12-01-00090 и № 13-07-00070.

Efficient approximation algorithms with performance guarantees for some discrete optimization problems in analysis and recognition of sequences

Kel'manov Alexander^{1,2}

kelm@math.nsc.ru

¹Novosibirsk, Sobolev Institute of Mathematics, SB RAS

²Novosibirsk, Novosibirsk State University

This report is on some discrete optimization problems induced by actual issues, in particular, in analysis and recognition of sequences. The purpose of the report is to overview new and recent results (obtained in Sobolev Institute of Mathematics) on studying computational complexity of these problems and on substantiating polynomial algorithms with performance guarantees for solving these problems. The considered problems have applications in approximation theory, computational geometry, and statistics. These problems are also important in a wide range of applications. For example, they arise in the medical and technical diagnostics, biometrics, electronic intelligence, in the aerospace monitoring, remote sensing of Earth and Ocean, in the criminal science and experimental data processing, and so on. The essence of these problems is to find some informationally important subsets and subsequences in the given noisy data arrays, to partition these data into subsets and subsequences, and to make a decision about objects generating these data.

The main results are of the following sort. Some new (previously unstudied), known (weakly studied), and open problems were investigated. Furthermore, some generalizations and special cases of classical partitioning problems and also some problems of searching for vectors subsets and subsequences in Euclidean space were studied. NP-hard and polynomial solvable cases of these discrete optimization problems are found. Exact polynomial and pseudopolynomial algorithms and also efficient approximation algorithms with performance guarantees for these problems are designed.

This research is funded by RFBR, grants 12-01-00090 and 13-07-00070.

Приближенный полиномиальный алгоритм для одной задачи разбиения последовательности

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

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

²Новосибирск, Новосибирский государственный университет

Рассматривается следующая NP-трудная в сильном смысле [1]

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

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

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

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

на элементы набора \mathcal{M} .

Задача моделирует актуальную проблему помехоустойчивого анализа сигналов и заключается в разбиении конечной последовательности векторов евклидова пространства на два кластера по критерию минимума суммы квадратов расстояний от элементов кластеров до их центров. Центр первого кластера является оптимизируемой величиной и определяется как среднее значение по всем векторам, образующим этот кластер. Центр второго кластера фиксирован в начале координат. Предложен 2-приближенный полиномиальный алгоритм решения задачи. Алгоритм реализует схему динамического программирования и выполняется за время $\mathcal{O}(N^2(T_{\max} - T_{\min} + q))$. Поскольку $T_{\max} - T_{\min} < N$, его трудоемкость можно оценить как $\mathcal{O}(N^2(N + q))$ в общем случае, когда $T_{\max} \neq T_{\min}$, и как $\mathcal{O}(N^2q)$ в случае $T_{\max} = T_{\min}$.

Работа поддержана грантами РФФИ № 12-01-00090 и № 13-07-00070.

[1] Кельманов А. В., Хамидуллин С. А. Приближенный полиномиальный алгоритм для одной задачи бикластеризации последовательности // Журн. вычисл. математики и мат. физики, 2014 (в печать).

Efficient approximation algorithm for a sequence partitioning problem

Kel'manov Alexander^{1,2}

kelm@math.nsc.ru

Khamidullin Sergey^{1*}

kham@math.nsc.ru

¹Novosibirsk, Sobolev Institute of Mathematics, SB RAS

²Novosibirsk, Novosibirsk State University

It is considered the following strongly NP-hard [1]

Problem 1-MSSC-S-NF. Given a sequence $\mathcal{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)$ of vectors from \mathbb{R}^q , and some positive integer numbers T_{\min} and T_{\max} . Find a subset $\mathcal{M} = \{n_1, \dots, n_M\} \subseteq \mathcal{N} = \{1, \dots, N\}$ such that

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

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

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

on the elements of \mathcal{M} .

The considered problem is actual, in particular, in the noise-proof analysis of signals.

The problem is to find a partition of a given Euclidean vectors sequence into two clusters minimizing the sum of squared distances from the clusters elements to their centers. The center of the first cluster is defined as the mean values of all vectors in a cluster. The center of the second cluster is given in advance and is equal to 0.

In this work, a 2-approximation efficient algorithm is presented. This algorithm implements a dynamic programming scheme and runs in time $O(N^2(T_{\max} - T_{\min} + q))$ where $T_{\max} - T_{\min} < N$. Therefore, algorithm is polynomial and it finds the solution in $O(N^2(N + q))$ time in the general case when $T_{\max} \neq T_{\min}$ and in $O(qN^2)$ time in the special case when $T_{\min} = T_{\max}$.

This research is funded by RFBR, grants 12-01-00090 and 13-07-00070.

- [1] Kel'manov A., Khamidullin S. 2014 (in press). An efficient approximation algorithm for a sequence bi-partitioning problem. *Zh. Vychisl. Mat. Mat. Fiz.* [Comput. Math. Math. Phys.].

Точный псевдополиномиальный алгоритм для одной задачи двухкластерного разбиения множества векторов

Кельманов Александр Васильевич kelm@math.nsc.ru
*Хандеев Владимир Ильич** vladimir.handeev@gmail.com

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

В докладе рассматривается следующая NP-трудная в сильном смысле

Задача 1-MSSC-F. Дано: множество $\mathcal{Y} = \{y_1, \dots, y_N\}$ векторов из \mathbb{R}^q и натуральное число M . Найти: разбиение множества \mathcal{Y} на два кластера \mathcal{C} и $\mathcal{Y} \setminus \mathcal{C}$ такое, что

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

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

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

Показано, что задача разрешима за время $\mathcal{O}(q^2 N^{2q})$, полиномиальное при фиксированной размерности q пространства. Для подслучая этого случая (когда компоненты векторов целочисленны) обоснован точный псевдополиномиальный алгоритм с временной сложностью $\mathcal{O}(N(MD)^q)$, где D — максимальное абсолютное значение координат векторов входного множества; алгоритм эффективнее известного полиномиального алгоритма при $MD < N^{2-1/q}$ [1].

Работа поддержана грантами РФФИ № 12-01-00090 и № 13-07-00070.

- [1] Кельманов А. В., Хандеев В. И. Точный псевдополиномиальный алгоритм для одной задачи двухкластерного разбиения множества векторов // Дискрет. анализ и исслед. операций, 2014. Т. 21. № 6. С. 3–13.

An exact pseudopolynomial algorithm for a vectors set bi-partitioning problem

Alexander Kel'manov

kelm@math.nsc.ru

Vladimir Khandeev*

vladimir.khandeev@gmail.com

Novosibirsk, Novosibirsk State University, Sobolev Institute of
Mathematics SB RAS

In this paper, it is considered the following strongly NP-hard

Problem 1-MSSC-F. *Given a set $\mathcal{Y} = \{y_1, \dots, y_N\}$ of vectors from \mathbb{R}^q and a positive integer number M . Find a partition of \mathcal{Y} into clusters \mathcal{C} and $\mathcal{Y} \setminus \mathcal{C}$ such that*

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

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

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

It is shown that the problem is solvable in $\mathcal{O}(q^2 N^{2q})$ time. Then, the problem is polynomially solvable if dimension q of space is fixed.

In addition, a pseudopolynomial algorithm, which finds an optimal solution in the case of integer values of the components of the vectors in the input set and fixed space dimension is presented. The running time of the algorithm is $\mathcal{O}(qN(2MD + 1)^q)$ where D is the maximum absolute coordinate value of the vectors in the input set [1].

This research is funded by RFBR, grants 12-01-00090 and 13-07-00070.

- [1] Kel'manov A. V., Khandeev V. I. 2014. An exact pseudopolynomial algorithm for a vectors set bi-partitioning problem. *Diskret. Anal. Issled. Oper.* 21(6):3–13.

Идентификация кристаллических решеток на основе оценивания параметров элементарных ячеек Браве и Вигнера–Зейтца

Кири Дмитрий Викторович^{1,2*}

limitk@mail.ru

*Куприянов Александр Викторович*¹

alexkupr@gmail.com

¹ Самара, Институт систем обработки изображений Российской академии наук

² Самара, Самарский государственный аэрокосмический университет имени академика С. П. Королёва (национальный исследовательский университет)

Рассматривается задача идентификации кристаллических решеток. Для решения этой задачи были разработаны два метода параметрической идентификации. В основе методов лежит вычисление меры схожести оцененных параметров исследуемой решетки с эталонными параметрами.

Первый метод оценивает шесть параметров элементарной ячейки Браве (длины трех сторон и величины трех углов между ними) с помощью поворотов решетки вокруг координатных осей. Второй метод оценивает объем ячейки Вигнера–Зейтца с помощью метода Монте-Карло. В обоих методах введены нормированные меры схожести: мера схожести сторон и мера схожести углов для элементарных ячеек Браве, мера схожести объемов для ячеек Вигнера–Зейтца.

С использованием методов параметрической идентификации было проведено исследование возможности структурной идентификации кристаллических решеток. Полученные результаты показали, что совместное применение двух разработанных методов обеспечивает точность структурной идентификации решеток четырех кристаллических систем (триклинной, моноклинной, тригональной и гексагональной) на уровне более 95% [1].

Работа поддержана грантами РФФИ: № 12-01-00237-а, № 14-01-00369-а и № 14-07-97040-р_поволжье_а.

- [1] *Кири Д. В.* Идентификация кристаллических решеток по координатам их узлов в трехмерном пространстве // Распознавание образов и анализ изображений: Успехи в области математической теории и приложений, 2014 (в печати). № 4.

Crystal lattice identification based on the estimation of Bravais and Wigner–Seitz unit cell parameters

Kirsh Dmitry^{1,2*}

limitk@mail.ru

*Kupriyanov Alexander*¹

alexkupr@gmail.com

¹Samara, Image Processing Systems Institute of the Russian Academy of Sciences

²Samara, Samara State Aerospace University named after S. P. Korolyov (national research university)

The problem of crystal lattice identification is considered. Two parametric identification methods were developed to solve this problem. The methods are based on the calculation of similarity measure between estimated parameters of the analyzed lattice and sample parameters.

The first method estimates six parameters of Bravais unit cell (lengths of the three edges and values of the three angles between them) by rotating the crystal lattice around the coordinate axes. The second method estimates volume of Wigner–Seitz cell with the use of the Monte-Carlo method. Both methods include the normalized similarity measures: edge similarity measure and angle similarity measure for Bravais unit cells and volume similarity measure for Wigner–Seitz cells.

The possibility of structural identification of crystal lattices was investigated using the parametric identification methods. The results demonstrated that the combined usage of the two developed methods ensure the structural identification accuracy for four lattice systems (triclinic, monoclinic, rhombohedral, and hexagonal) at over 95% [1].

This research is funded by RFBR, grants 12-01-00237-a, 14-01-00369-a, and 14-07-97040-p_povolzh'e_a.

- [1] Kirsh D. 2014 (in press). Crystal lattice identification by coordinates of their nodes in three dimensional space. *Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications 4*.

Аппроксимируемость геометрической задачи о нескольких коммивояжерах

*Хачай Михаил Юрьевич**

mkhachay@imm.uran.ru

Незнахина Екатерина Дмитриевна

eneznakhina@yandex.ru

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

Исследуется задача k -МНС об оптимальном покрытии полного взвешенного графа фиксированным числом вершинно непересекающихся циклов. Критерий оптимизации аддитивен и равен суммарному весу дуг, входящих в построенные маршруты. Задача обладает рядом практически важных приложений, в частности, в таксономии, и является естественным обобщением известной труднорешаемой задачи коммивояжера (TSP) (совпадающей с k -МНС при $k = 1$), исследование вопросов аппроксимируемости которой является одним из основных направлений современной теории вычислительной сложности. Известно, что в общем случае задача TSP полиномиально не аппроксимируема с произвольной точностью 2^n (при условии $P \neq NP$), однако уже в метрической постановке принадлежит классу Арх. Как показано С. Аророй, евклидова задача TSP (в которой вершины графа являются точками в конечномерном числовом пространстве, а веса дуг определяют расстояния между ними) обладает полиномиальной приближенной схемой (polynomial-time approximation scheme, PTAS), т. е. в наиболее часто встречающемся на практике частном случае задача может быть решена приближенно с произвольной наперед заданной точностью. Показано, что задача k -МНС при произвольном фиксированном $k \geq 1$ обладает аналогичными свойствами, а именно: задача NP-трудна в сильном смысле и сохраняет труднорешаемость в метрической и евклидовой постановках; задача неаппроксимируема в общем случае; метрическая задача k -МНС обладает 2-приближенным алгоритмом; для евклидовой задачи предложена и обоснована PTAS [1].

Работа поддержана грантом РФФ № 14-11-00109.

- [1] *Хачай М. Ю., Незнахина Е. Д.* Полиномиальная приближенная схема для задачи о двух коммивояжерах на плоскости // Труды института математики и механики, 2015 (в печати). Екатеринбург: УрО РАН.

Polynomial-time approximability of geometric multiple traveling salesmen problem

*Khachay Michael**

mkhachay@imm.uran.ru

Neznakhina Ekaterina

eneznakhina@yandex.ru

Ekaterinburg, Krasovsky Institute of Mathematics and Mechanics UB
RAS

For a complete weighted graph, the optimal covering problem by k vertex-disjoint cycles (k -MHC) is studied. Optimization criterion is an additive one, it is required to find a cover having the minimum total weight of the belonging arcs. The problem has multiple applications, e. g., in cluster analysis, and naturally generalizes the well-known travelling salesmen problem (TSP) (TSP is 1-MHC).

The general TSP problem is NP-hard and cannot be approximated within 2^n (unless $P = NP$). Meanwhile, even metric TSP belongs to Apx. Arora has shown that euclidean TSP (where vertices are the points in some finite-dimensional euclidean space and weights are the euclidean distances among them) has polynomial-time approximation scheme (PTAS). That is, due to this shining result, the most popular subclass of TSP problem can be approximated efficiently in polynomial time within any given accuracy.

The obtained results are the following. It is proved that k -MHC is strongly NP-hard for any fixed $k \geq 1$ and remains intractable in metric and euclidean cases. It is proposed 2-approximation polynomial-time algorithm for metric k -MHC and PTAS for euclidean k -MHC [1].

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

- [1] Khachay M., Neznakhina E. 2015 (in press). Polynomial-time approximation scheme for two salesmen problem in the plane. *Proceedings of the Steklov Institute of Mathematics*. Moscow: MAIK.

Тернарная машинная арифметика: история, проблемы, приложения

Чернов Владимир Михайлович^{1,2}

vche@smr.ru

¹ Самара, Самарский государственный аэрокосмический университет имени академика С. П. Королёва (Национальный исследовательский университет)

² Самара, Институт систем обработки изображений РАН

Представлен обзор результатов, относящихся к проблемам тернарной машинной арифметики: история, методы, идеи, перспективы и приложения [1].

- [1] *Богданов П. С., Чернов В. М.* Классификация тернарных квазиканонических систем счисления в мнимых квадратичных полях и их приложение // Компьютерная оптика, 2014. Т. 38. № 1. С. 139–147.

Ternary computer arithmetic: History, problems, applications

Chernov Vladimir^{1,2}

vche@smr.ru

¹Samara, Samara State Aerospace University

²Samara, Image Processing System Institute of RAS

An overview of the results referring to the problems of ternary computer arithmetic: history, techniques, ideas, perspectives and applications, has been provided [1].

- [1] Bogdanov P., Chernov V. 2013. Classification of ternary quasicanonical number systems in imaginary quadratic fields and their applications. *Computer Optics* 38(1):139–147. [In Russian.]

Использование модели социальной сети с сообществами пользователей для распределенной генерации случайных социальных графов

*Чихрадзе Кирилл Константинович** chykhradze@ispras.ru

Коршунов Антон Викторович korshunov@ispras.ru

Кузюрин Николай Николаевич nnkuz@ispras.ru

Москва, ИСП РАН

Онлайновые социальные сети на сегодняшний день являются одним из наиболее популярных типов социальных сервисов в Интернете, и их размеры достигают сотен миллионов пользователей. Для исследования структурных свойств таких сетей применяются алгоритмы поиска сообществ пользователей. Учитывая большое количество пользователей, необходимы алгоритмы, чья эффективность доказана на больших социальных графах. Стандартный способ оценки эффективности таких алгоритмов — тестирование на случайных графах со структурой сообществ.

Цель настоящей работы — разработка моделей, методов и программных средств для генерации случайных социальных графов с известной структурой пересекающихся сообществ пользователей.

В работе описан распределенный алгоритм, позволяющий создавать случайные графы с сообществами, структура которых удовлетворяет основным свойствам реальных социальных сетей.

В ходе тестирования программной реализации алгоритма в рамках фреймворка Apache Spark установлено, что алгоритм обладает линейной масштабируемостью. При этом графы, созданные согласно данному алгоритму, удовлетворяют некоторым важным свойствам социальных сетей (размер сообщества и количество сообществ у пользователя распределены по степенному закону, повышенная плотность ребер в пересечениях сообществ и др.). Также были протестированы различные алгоритмы поиска сообществ [1].

- [1] *Чихрадзе К. К., Коршунов А. В., Кузюрин Н. Н.* Использование модели социальной сети с сообществами пользователей для распределенной генерации случайных социальных графов // Машинное обучение и анализ данных, 2014. Т. 1. № 8. С. 1027–1047. <http://jmla.org/papers/doc/2014/no8/Chykhradze2014Communities.pdf>.

On a model of social network with user communities for distributed generation of random social graphs

*Chykhradze Kyrylo**

chykhradze@ispras.ru

Korshunov Anton

korshunov@ispras.ru

Kuzyurin Nikolay

nnkuz@ispras.ru

Moscow, ISP RAS

Today, online social networks are one of the most popular types of social services on the Internet, and their sizes reach hundreds of millions of users. User community detection algorithms are used to investigate structural properties of such networks. However, large number of users demands algorithms whose effectiveness is proven on large graphs. A standard method for assessing the effectiveness of such algorithms is using random graphs with community structure.

The goal of the present work is to develop models, methods, and software for generating random social graphs with known overlapping community structure of users.

The paper describes a distributed algorithm which allows to create random graphs with community structure that satisfies the basic properties of real social networks.

Evaluation of the distributed Apache Spark implementation showed that the algorithm has linear scalability. The graphs generated according to the algorithm satisfy some of the basic properties of social networks (e. g., power-law distribution of user-community membership and community size, dense intersections of communities, etc.). Various community detection algorithms are also evaluated [1].

- [1] Chykhradze K., Korshunov A., Kuzyurin N. 2014. On a model of social network with user communities for distributed generation of random social graphs. *Machine Learning and Data Analysis* 1(8):1027–1047. <http://jmla.org/papers/doc/2014/no8/Chykhradze2014Communities.pdf>.

Адаптивная процедура обобщенного сглаживания изображений на основе статистического подхода

*Грачева Инесса Александровна**

gia1509@mail.ru

Копылов Андрей Валериевич

And.Kopylov@gmail.com

Красоткина Ольга Вячеславовна

ko180177@yandex.ru

Тула, Тульский государственный университет

Сглаживание размытых локальных оценок имеет своей целью максимальное подавление локальных возмущений без существенного ущерба точности восстановления значений скрытого параметра нестационарного сигнала или изображения, поскольку в основе такого подхода к анализу сигналов или изображений лежит априорное предположение, что результат обработки является достаточно гладким и умеренная степень сглаживания в процессе оценивания не может его существенно исказить. Однако при обработке часто возникает ситуация, когда такое априорное предположение не является вполне адекватным природе анализируемого изображения и необходимо допустить наличие явных разрывов в подлежащей восстановлению скрытой нестационарной модели, замаскированных шумом в анализируемых данных.

В работе рассматривается обобщение байесовского подхода к анализу изображений, включающее автоматическую адаптацию параметров сглаживания к наблюдаемым данным, позволяющую сохранить существенные локальные особенности изображения, в частности границы объектов. Предлагаемая гамма-нормальная модель изображения и ожидаемого результата обработки, рассматриваемых как двухкомпонентное случайное поле, позволила разработать параметрический адаптивный алгоритм сглаживания изображений. Предложенный алгоритм, в отличие от существующих, является простым в настройке и обладает линейной вычислительной сложностью относительно количества элементов изображения.

Работа поддержана грантом РФФИ № 13-07-00529.

- [1] *Грачева И. А., Копылов А. В.* Адаптивный параметрический алгоритм обработки изображений // Машинное обучение и анализ данных, 2014 (в печати).

Adaptive procedure for generalized smoothing of images on the basis of statistical approach

*Gracheva Inessa**

gia1509@mail.ru

Kopylov Andrey

And.Kopylov@gmail.com

Krasotkina Olga

ko180177@yandex.ru

Tula, Tula State University

Smoothing of fuzzy local estimates is aimed at the maximum suppression of the local peculiarities without significant loss of accuracy of the recovered hidden parameters of a nonstationary signal or image. The basis for such approach to the analysis of signals or images is the *a priori* assumption that the sought for result of processing is smooth enough and a moderate degree of smoothing cannot significantly distort it. However, it is common situation, when such *a priori* assumption is not quite adequate to the nature of the analyzed image, and one should obviously introduce the possibility of abrupt changes into the hidden nonstationary model, disguised by noise in analyzed data.

In this paper, a generalization of the Bayesian approach to the image analysis, which involves automatic adjustment of the changing smoothness parameters to the observable data, is considered. Such generalized approach allows one to preserve substantial local image features, in particular, edges of objects. Proposed gamma-normal model of the image and the expected result of processing, considered as a two-component random field, allowed to develop parametric adaptive algorithm for image smoothing. In contrast to the existing algorithms, the algorithm being proposed is simple in tuning and has linear computation complexity with respect to the number of image elements.

This research is funded by RFBR, grant 13-07-00529.

- [1] Gracheva I., Kopylov A. 2014 (in press). Adaptive parametric image processing algorithms. *Machine Learning and Data Analysis*.

Симметризация точек изображения, заданных статистическими выборками

*Каркищенко Александр Николаевич** karkishalex@gmail.com

Мнухин Валерий Борисович mnukhin.valeriy@mail.ru

Таганрог, Южный федеральный университет

Работа посвящена построению симметричных конфигураций точек изображения по заданным, но не обладающим этим свойством точкам, про которые известно, что они должны быть симметричны. Данная задача называется задачей симметризации. В предыдущих работах авторов рассматривались методы ее решения при условии, что каждая точка задана некоторой оценкой своего положения, которое может не совпадать с истинным. Были предложены решения для случаев отражательной и вращательной симметрии с разной степенью априорной неопределенности. Оптимальность получаемых решений выражалась в том, что симметризация достигалась минимальным уклонением «симметризованных» точек от первоначальных положений. В отличие от предшествующих публикаций, в данной статье рассматривается более общий случай, когда точки задаются статистическими выборками координат. При этом требуется, соответственно, «статистическое» понимание оптимальности решений. Предлагаются методы построения по статистическим выборкам гарантированно симметричных конфигураций, оптимальных в том смысле, что они обладают наибольшей плотностью вероятности появления. Показано, что эти задачи сводятся к стандартной постановке задачи минимизации в пространстве с метрикой Махаланобиса. Рассматриваются задачи «статистической» симметризации относительно отражательной и вращательной симметрий. Предлагается также нормированная мера симметричности исходных статистических данных. Для иллюстрации методов симметризации приводятся результаты моделирования [1].

Работа поддержана грантами РФФИ № 13-07-00327-а и № 13-07-13112-офи_м_РЖД.

- [1] Каркищенко А. Н., Мнухин В. Б. Симметризация точек изображения, заданных статистическими выборками // Машинное обучение и анализ данных, 2014. Т. 1. № 7. С. 920–935. <http://jmla.org/papers/doc/2014/no7/Karkishchenko2014Symmetrization.pdf>.

Symmetrization of the image points defined by statistical sampling

*Karkishchenko Alexander**

karkishalex@gmail.com

Mnukhin Valeriy

mnukhin.valeriy@mail.ru

Taganrog, Southern Federal University

The paper is devoted to the methods of constructing symmetrical configurations of characteristic points in an image calculated from a certain set of given points that do not possess this property, but about which it is known *a priori* that they should be symmetrical. This problem has many applications and is called the problem of symmetrization. In their previous studies, the authors examined the methods for its solution provided that each point is given with some estimate of its position, which may not coincide with the true one. Such a situation is typical when processing images. Solutions have been proposed for the cases of reflectional and rotational symmetry with varying degrees of *a priori* uncertainty. Optimality of the solutions is in the fact that symmetrization is achieved in a minimum deviation of “symmetrized” points from the original locations. Here, the more general case when the points are specified by statistical samples is considered. This case requires the “statistical” understanding of the optimality. The methods proposed in this paper, being based on statistical samples, are available for constructing symmetrical configurations which are optimal in the sense that they have the highest probability density of appearance. The paper shows that the problem can be reduced to the standard setting of the minimization problem in a space with Mahalanobis metric. The problems of “statistical” symmetrization of points related to both reflectional and rotational symmetry are considered here. Also, normalized measure is proposed to assess the degree of symmetry of the source statistics. To illustrate the methods of symmetrization, the results of modeling are given in the paper [1].

This research is funded by RFBR, grants 13-07-00327 and 13-07-13112-ofi_m_RZhD.

- [1] Karkishchenko A., Mnukhin V. 2014. Symmetrization of the image points defined by statistical sampling. *Machine Learning and Data Analysis* 1(7):920–935. <http://jmla.org/papers/doc/2014/no7/Karkishchenko2014Symmetrization.pdf>.

Разработка трехмерного сканера человеческого тела для проекта виртуальной примерочной с использованием сенсоров Kinect

Ларин Александр Олегович

ekzebox@gmail.com

Москва, Московский физико-технический институт

Появление на потребительском рынке дешевых устройств с RGB-D камерой: MS Kinect, ASUS Xtion Pro, PrimeSense, Intel и др., — позволяющих получать помимо обычного цветового изображения трехмерное (3D) облако точек, представляющих объект, породило большую волну публикаций и технических решений, посвященных разработке недорогого 3D сканера для личных нужд и проектов с ограниченным бюджетом.

К сожалению, большая часть из этих работ и решений являются либо полностью закрытыми и очень дорогими, либо бесплатными и открытыми, но сложными в поддержке и позволяющими получать низкое качество результата, а публикации имеют разрозненный характер, охватывают очень узкий круг задач и не дают четкого и законченного представления об алгоритмах, необходимых для разработки полностью автоматического программного обеспечения. Разработка 3D сканера даже на основе данных с RGB-D камеры, даже в условиях, казалось бы, большого количества существующего материала, остается достаточно сложной задачей, охватывающей широкий спектр задач из областей Surface Reconstruction и Computer Vision: устранение шумов в 3D облаке точек и его сглаживание, восстановление 3D поверхности в реальном времени, текстурирование полученной модели с применением морфологии и фильтрации.

Данная работа представляет собой содержательное описание набора технических решений и алгоритмов, позволяющих получать текстурированную 3D модель человека приемлемого качества, достаточного для решения широкого спектра прикладных задач, например приложений виртуальной реальности или в качестве основы проекта виртуальной примерочной [1].

- [1] *Ларин А. О.* Разработка 3D сканера человеческого тела для проекта виртуальной примерочной с использованием сенсоров Kinect // Машинное обучение и анализ данных, 2014 (в печати).

The development of three-dimensional human body scanner for virtual fitting room using Kinect

Larin Alexander

ekzebox@gmail.com

Moscow, Moscow Institute of Physics and Technology

The market appearance of cheap consumer devices with RGB-D camera such as MS Kinect, ASUS Xtion Pro, PrimeSense, Intel, etc., which, besides the usual color image, give three-dimensional (3D) point cloud representation of object, has generated a great wave of publications and technical solutions devoted to the development of low-cost 3D scanner for personal needs and projects with a limited budget.

Unfortunately, most of these works and solutions are either completely closed and very expensive, or free and open, but difficult to support and bring poor results. The publications are fragmented, cover a very narrow range of tasks, and do not provide clear and complete idea of the algorithms needed to develop a fully automated software. Even in the presence of seemingly large amount of existing material, the development of 3D scanner based on data (received) from RGB-D camera, remains a rather difficult task, covering a wide range of problems in areas of Surface Reconstruction and Computer Vision like elimination of noise in a 3D point cloud and its smoothing, real-time 3D surface reconstruction, and texturing of the obtained pattern using morphology and filtration.

This work presents extended description of set of technical solutions and algorithms, which allows receiving a textured 3D model of a person of acceptable quality, sufficient for solving a wide range of problems, such as the developing of virtual reality applications or using as a base project virtual fitting [1].

- [1] Larin A. 2014 (in press). The development of 3D human body scanner for virtual fitting room using Kinect. *Machine Learning and Data Analysis*.

Интеллектуальные возможности гипертрейс-преобразования: конструирование признаков с заданными свойствами

*Федотов Николай Гаврилович*¹ fedotov@pnzgu.ru

Сёмов Алексей Александрович^{1*} matematik_aleksey@mail.ru

*Моисеев Александр Владимирович*² moigus@mail.ru

¹Пенза, Пензенский государственный университет

²Пенза, Пензенский государственный технологический университет

В данной работе предлагается новый подход к распознаванию 3D объектов, основанный на методах стохастической геометрии и функционального анализа, который обладает возможностями интеллектуального анализа данных. Признаки имеют гипертриплетную композиционную структуру, способствующую простой реализации алгоритма и конструированию большого числа признаков. Строгая модель позволяет строить признаки аналитически, описывая каждый класс объектов и их особенности.

3D трейс преобразование позволяет создавать инвариантное описание пространственного объекта, которое более устойчиво к искажениям и шумам, чем получаемое в результате процедуры нормализации объекта. Возможность регулировать свойства признаков повышает интеллектуальные возможности преобразования, что является его преимуществом. В работе приводятся примеры гипертриплетных признаков, имеющих заданные свойства.

В статье анализируется роль функционалов, входящих в структуру гипертриплетного признака. Описываются расширенные возможности 3D трейс преобразования, например, извлечение в той же технике сканирования информации о пространственном положении и ориентации объекта. Приводится описание способов интеллектуального анализа 3D изображений, например, высокоуровневая предобработка 3D изображений, выполненная в той же технике сканирования, что и формирование признаков.

Работа поддержана грантом РФФИ № 12-07-00501.

- [1] *Федотов Н. Г., Сёмов А. А., Моисеев А. В.* Интеллектуальные возможности гипертрейс-преобразования: конструирование признаков с заданными свойствами // Машинное обучение и анализ данных, 2014 (в печати).

Intelligent capabilities hypertrace transform: constructing features with predetermined properties

*Fedotov Nikolay*¹

fedotov@pnzgu.ru

Syemov Aleksey^{1*}

mathematik_aleksey@mail.ru

*Moiseev Alexandr*²

moigus@mail.ru

¹Penza, Penza State University

²Penza, Penza State Technological University

In this article a new approach to the 3D objects' recognition based on modern methods of stochastic geometry and functional analysis is proposed. This method has many advantages and data mining capabilities. Thus, features have hypertriplet composite structure, which provide not only easy machine implementation of this algorithm, but construction of a large number of features. Due to building a rigorous mathematical model, the analyst can construct analytical and not intuitive features, describing each object class and their features (in particular, constructing geometric features).

3D trace transform allows to create invariant description of spatial object, which is more resistant to distortion and coordinate noise than the description obtained as a result of the object normalization procedure. Possibility of regulating constructed features' properties significantly increase intellectual capabilities of 3D trace transform that is undoubtedly its advantage. Proof developed theory and the mathematical model is variety constructed theoretical examples of hypertriplet features having described particular properties.

In the article role of functional included in compositional structure of hypertriplet feature is analyzed. Extended possibilities of 3D trace transform, in particular, extracting in the same scanning technique the information about the spatial position and orientation of three-dimensional object are described. Description of many ways of 3D image mining is proposed. For example, one of the intellectual abilities of the proposed method is a high-level preprocessing, processing and post-processing of 3D images in one scanning technique.

- [1] Fedotov N., Syemov A., Moiseev A. 2014 (in press). Intelligent capabilities hypertrace transform: constructing features with predetermined properties. *Machine Learning and Data Analysis*.

Метод полуавтоматической контекстной разметки сцены

*Егоров Антон Иванович**

aegorov@gosniias.ru

Вишняков Борис Ваисович

bvishnyakov@gosniias.ru

Москва, ФГУП ГосНИИАС

Контекстная разметка сцены является одной из наиболее актуальных задач для широкого спектра приложений, работающих с изображениями или видеоданными. Одной из основных областей применения для контекстной разметки сцены является видеонаблюдение. Использование такой контекстной информации о сцене, как расположение дорог, тротуаров, зданий, газона и пр., позволяет значительно увеличить вероятности обнаружения объектов интереса и уменьшить вероятности ложных срабатываний как алгоритмов анализа движения по видеопоследовательности, так и алгоритмов поиска объектов интереса, применяемых на отдельных изображениях и основанных на машинном обучении.

Предлагаемый метод предполагает полуавтоматическую контекстную разметку сцены, организованную по итерационному принципу, что освобождает оператора от рутинной ручной разметки каждого пикселя изображения, однако оставляет ему возможность вносить уточнения в сгенерированную разметку в любом масштабе, начиная от уточнения отдельных пикселей, заканчивая полной переразметкой «с нуля».

Данный метод принимает на вход некоторую неточную разметку объектов на сцене. Данная разметка выполняется оператором вручную и соотносится с полученными после сегментации регионами. Соответствие определяется по мажоритарному принципу, исходя из количества пикселей каждого из сегмента, размеченного оператором [1].

Предложенный алгоритм был протестирован на изображениях из общедоступной базы PETS.

Работа выполнена при поддержке РФФИ, грант 12-07-00798-а.

- [1] *Егоров А. И.* Метод полуавтоматической контекстной разметки сцены // Вестник компьютерных и информационных технологий, 2014 (в печати).

Method of semiautomatic contextual scene markup

Egorov Anton *

aegorov@gosniias.ru

Vishnyakov Boris

bvishnyakov@gosniias.ru

Moscow, GosNIIAS

Contextual scene markup is one of the most actual problems for a wide range of image and videoprocessing applications. Using such information as location of road, sidewalks, buildings, lawns, etc. on the scene can significantly increase the probability of successful object detection and reduce the probability of false positives.

In this paper, a method of contextual scene markup is suggested. Based on a prior scene markup from a user, the method allows identifying the type of areas or objects presented on the scene. Image preparing stage includes the following steps: smoothing, brightness gradients extraction using Sobel operator, nonmaximums suppression and morphological restoration of contour boundaries breaks. The process of initial image segmentation and its further elaboration by absorbing small segments are described. Finally, is proposed an algorithm for matching segments and initial scene markup by operator is proposed.

The method is organized by the iterative principle, so that the operator can make any number of refinements after the algorithm completion and if necessary run it again. Otherwise, contextual markup becomes final and goes to the surveillance algorithms [1].

The results of the proposed methods are obtained on test video database PETS.

This research is funded by RFBR, grant 12-07-00798-a.

- [1] Egorov A. 2014 (in press). Method of semi-automatic contextual scene markup. *Herald of Computer and Information Technologies*.

Сравнение изображений картин по информативным фрагментам

Мурашов Дмитрий Михайлович^{1*}

d_murashov@mail.ru

*Березин Алексей Владимирович*²

berezin_alex@mail.ru

*Иванова Екатерина Юрьевна*²

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

²Москва, Государственный исторический музей

Рассматривается задача сравнения изображений картин по информативным фрагментам для целей атрибуции произведений живописи. Используемые текстурные признаки описывают фактуру картины и являются характеристикой стиля живописи художника. В качестве признаков применялись гистограммы локальной ориентации хребтов полутонового рельефа изображения и локальной ориентации простых окрестностей. Разработана процедура извлечения признаков. Для сравнения информативных фрагментов применяется теоретико-информационная мера различия на основе дивергенции Кульбака–Лейблера. Предложена методика сравнения изображений картин по значениям меры различия фрагментов. Методика тестировалась на изображениях портретов, написанных в XVIII–XIX вв. Сравнение картин проводилось по трем однотипным фрагментам, выделенным на каждом изображении. Результаты экспериментов показали, что различия картин одного автора существенно меньше различий между картинами разных авторов. Предложенная методика сравнения изображений может быть использована в качестве одного из видов технико-технологических исследований картин. Дальнейшие исследования будут направлены на расширение признакового описания фрагментов картин, разработку методов поиска информативных фрагментов, увеличение базы изображений и тестирование разработанной методики на большем объеме данных [1].

Работа поддержана грантом РФФИ № 12-07-00668.

- [1] *Мурашов Д. М., Березин А. В., Иванова Е. Ю.* Сравнение изображений картин по информативным фрагментам // Машинное обучение и анализ данных, 2014. Т. 1. № 8. С. 941–948. <http://jmla.org/papers/doc/2014/no8/Murashov2014Comparing.pdf>.

Comparing images of paintings using informative fragments

Murashov Dmitry^{1*}

d_murashov@mail.ru

*Berezin Alexey*²

berezin.alex@mail.ru

*Ivanova Ekaterina*²

¹Moscow, Dorodnicyn Computing Centre of RAS

²Moscow, State Historical Museum

The problem of comparing images for the purpose of attribution of fine-art paintings is considered. Features that are used in this work describe texture of a painting and characterize the artistic style of a painter. The feature space is composed of a histogram of orientation angles of grayscale image ridges and a histogram of simple neighborhood orientation based on the local structure tensor. A procedure for feature extraction is developed. An information-theoretical dissimilarity measure based on Kullback–Leibler divergence is used for comparing informative image fragments. A method for comparing images of paintings using dissimilarity measure values between fragments is proposed. The method is tested on images of portraits created in XVIII–XIX centuries. The paintings are compared using three fragments segmented in a particular image. The results of the experiments showed that the difference between portraits painted by the same artist is substantially smaller than one between portraits painted by different authors. The proposed technique may be used as a part of techno-technological description of fine art paintings for attribution. The future research will be aimed at extending feature space, developing a technique for locating informative fragments, augmenting image dataset, and testing the developed method on augmented dataset [1].

This research is funded by RFBR, grant 12-07-00668.

- [1] Murashov D., Berezin A., Ivanova Y. 2014. Composing feature description of paintings texture. *Machine Learning and Data Analysis* 1(8):941–948. <http://jmla.org/papers/doc/2014/no8/Murashov2014Comparing.pdf>.

Новый метод реального времени для контекстного описания изображений и его применения к навигации роботов и интеллектуальному управлению

Кий Константин Иванович

konst.i.kiy@gmail.com

Москва, Институт прикладной математики им. М. В. Ёжельдыша РАН

Применение компьютерного зрения и понимания изображений имеют критическую важность в робототехнических системах будущего. Важнейшее значение будет иметь способность находить предметы по словесным описаниям или заданным эталонам. Для этих целей предлагается новый метод сегментации и описания цветных изображений и даются его применения к поиску и отслеживанию ориентиров в реальном времени. Ключевым объектом подхода является структурный граф цветовых сгустков STG, который в сжатой форме описывает основную информацию об изображении. Важнейшая информация, необходимая для принятия решений, заключена в контрастных объектах на изображениях реального мира. Разработаны новые методы, которые позволяют выделять эти объекты на STG. Для этих целей определяются левые и правые ростки контрастных объектов (left-right germs of contrast objects, LRG) на STG (непрерывные левые и правые контрастные граничные кривые на STG). На множествах левых и правых ростков контрастных объектов задается структура двудольного графа LRG. Контрастные объекты на STG интерпретируются как связанные компоненты на LRG и как некоторые контрастные объекты на изображении. Указанная техника успешно применена для движения робота на ориентиры с зрительной обратной связью [1].

Работа поддержана грантами РФФИ № 13-08-01118 и № 13-07-00988.

- [1] *Kiy K. I.* A new real-time method of contextual image description and its application in robot navigation and intelligent control // Computer vision in advanced control systems using conventional and intelligent paradigms. Springer Book Ser. 8578. Berlin: Springer, 2014. Vol. 2. P. 107–130.

A new real-time method of contextual image description and its application in robot navigation and intelligent control

Kiy Konstantin

konst.i.kiy@gmail.com

Moscow, Keldysh Institute of Applied Mathematics of RAS

Computer vision and image understanding are of crucial importance in robotic systems of the future. The ability of finding objects based on verbal descriptions and templates will be of crucial importance. For these purposes, a new method of segmentation and description of color images is proposed and its applications to finding and tracking landmarks in real time are given. The structural graph of color bunches STG, which concisely describes the main information of the image, is the key point of the approach. The most important information about the image is described by its contrast objects. In this paper, new methods that are able to select these objects in the STG have been developed. For these purposes, left and right germs (LRG) of contrast objects in STG (continuous left and right contrast boundary curves in STG) are constructed. On the sets of LRG of contrast objects, the structure of a bipartite graph LRG is introduced. Contrast objects in STG are interpreted as connected components in LRG and as contrast objects in the image. This technique has been successfully applied to arrangement of the motion of an autonomous robot to landmarks with pure visual feedback [1].

This research is funded by RFBR, grants 13-08-01118 and 13-07-00988.

- [1] Kiy K. 2014. A new real-time method of contextual image description and its application in robot navigation and intelligent control. *Computer vision in advanced control systems using conventional and intelligent paradigms*. Springer Book ser. 8578. Berlin: Springer. 2:107–130.

Поиск структурных различий изображений

Корнилов Федор Андреевич

fakornilov@mail.ru

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

Рассматривается задача поиска структурных различий на одновременных космических снимках земной поверхности, под которыми понимаются появившиеся или исчезнувшие объекты местности. В качестве отправной точки служит морфологический анализ изображений Ю. П. Пытьева.

Вводится общая формулировка понятия структуры изображения и структурных различий. Предлагается ряд алгоритмов, решающих поставленную задачу для полутоновых изображений. Первый алгоритм основан на морфологическом проекторе, второй представляет собой его регуляризованный вариант. Третий и четвертый алгоритмы основаны на методе наименьших квадратов, а последний использует марковские случайные поля. Для первого алгоритма приводятся методика построения распределения яркости результирующего изображения в случае возмущения входных изображений аддитивным шумом и определяются формула оптимального порога и оценки ошибок первого и второго рода. Предлагается также вариант морфологического проектора для работы с цветными изображениями.

Формулируется вычислительная методика сравнения алгоритмов поиска структурных различий изображений и результаты ее применения к описанным методам с целью определения их оптимальных параметров, а также приводятся результаты сравнения качества работы предложенных алгоритмов между собой. Эксперименты на реальных данных показывают, что предложенные алгоритмы подходят для прикладного применения [1].

Работа выполнена в рамках программ фундаментальных исследований Президиума РАН при финансовой поддержке УрО РАН «Динамические системы и теория управления» (проект 12-П-1-1022) и «Информационные, управляющие и интеллектуальные технологии и системы» (проект 12-П-1-1023).

[1] *Корнилов Ф. А.* Поиск структурных различий изображений // Машинное обучение и анализ данных, 2014. Т. 1. № 7. С. 902–919. <http://jmla.org/papers/doc/2014/no7/Kornilov2014Detection.pdf>.

Structural differences detection on the images

Kornilov Fedor

fakornilov@mail.ru

Ekaterinburg, Institute of Mathematics and Mechanics named after
N. N. Krasovkiy, UB RAS

The article is devoted to the problem of structural differences detection at satellite images, taken at different times. Here, structural changes mean appeared or disappeared ground objects. The starting point of the investigation is morphological analysis of images theory by Yu. P. Pytiev.

In the work, the general notion of the image's structure and structural difference are introduced. Several algorithms for solving the problem for grayscale images are proposed. The first algorithm is based on the morphological projector. The second algorithm is the regularized version of the first one. The third and fourth algorithms are based on the least-square method. The last algorithm uses Markov random fields. For the first algorithm, in the case of image's distortion with additive noise, the method for computing the resulted image's values distribution is given. Also, the formula for optimal threshold and estimation of the I and II type errors are given. The morphological projector's modification for dealing with color images is also provided.

The computing method for comparison of the algorithms of the images' structural differences detection is formulated. The results of its applying to the considered algorithms for searching algorithms' optimal parameters and comparing algorithms are given. The experiments with real data demonstrate that the proposed algorithms are good enough for applications [1].

This work was performed as part of the Programs of the Presidium of the Russian Academy of Sciences with financial support Ural Branch of the Russian Academy of Sciences (project nos. 12-P-1-1022 and 12-P-1-1023).

- [1] Kornilov F.A. 2014. Structural differences detection on the images. *Machine Learning and Data Analysis* 1(7):902–919. <http://jmla.org/papers/doc/2014/no7/Kornilov2014Detection.pdf>.

Топологическая фильтрация для распознавания и анализа симметрии цифровых изображений

Каркищенко Александр Николаевич karkishalex@gmail.com
*Мнухин Валерий Борисович** mnukhin.valeriy@mail.ru

Таганрог, Южный федеральный университет

Ряд методов обработки и анализа цифровых изображений в частотной области основан на формальном переносе свойств непрерывного преобразования Фурье на дискретный случай. Это зачастую приводит к искажениям, называемым *алиасинг* или *наложение*.

В работе алиасинг рассматривается как следствие различной топологии частотных областей непрерывного и дискретного преобразований Фурье (ДПФ): сферы в непрерывном случае и тора в дискретном. Вводится оператор *обвертки тора*, отображающий функции на плоскости в функции на торе. Доказывается, что ДПФ цифрового изображения является дискретизацией обвертки тора непрерывным Фурье-образом кусочно-постоянной аппроксимации этого изображения. Это позволяет представить ДПФ суммой ряда, начальный член которого совпадает с дискретизацией непрерывного Фурье-образа, а последующие члены интерпретируются как влияние высокочастотных гармоник.

На основе полученных результатов вводятся «оптимальные» операторы вращения в дискретных частотной и пространственной областях. Предлагается алгоритм *топологической фильтрации* дискретного спектра, позволяющий повысить достоверность распознавания и совмещения цифровых изображений, а также анализа их симметрии в частотной области.

Работа поддержана грантами РФФИ № 13-07-00327-а и № 13-07-13112-офи_м_РЖД.

- [1] Каркищенко А. Н., Мнухин В. Б. Топологическая фильтрация для распознавания и анализа симметрии цифровых изображений, 2014. Т. 1. №8. С. 966–987. <http://jmla.org/papers/doc/2014/no8/Mnukhin2014Topological.pdf>.

Topological filtration for digital images recognition and symmetry analysis

Karkishchenko Alexander

karkishalex@gmail.com

*Mnukhin Valeriy**

mnukhin.valeriy@mail.ru

Taganrog, Southern Federal University, Russia

A number of methods for digital images processing (say, the Fourier–Mellin transform method) is based on the formal adaptation of the continuous Fourier transform properties to the discrete case. As a result, aliasing errors occur and it is important to control it.

In the paper, aliasing is considered as a topological effect, caused by different properties of frequency domains of continuous and discrete Fourier transforms (DFT). Indeed, the continuous frequency domain may be considered as a sphere but it is a torus in the discrete case.

The present authors define a *torus winding operator* that transforms functions in a plane (or a sphere) to functions on a torus. It is shown that the discrete spectrum of an image is the discretization of the winding of a torus by the continuous Fourier transform of the piecewise-constant approximation of the image. As a corollary, DFT is expanded into a series, whose initial term is the continuous Fourier transform discretization, and the terms are of higher harmonic effects. The results are applied to the problem of finding the “optimal” rotations in discrete frequency and spatial domains.

Another application of the stated results is a technique to reduce systematic errors in frequency domain-based methods for digital images recognition, registration, and symmetry analysis. It is based on the continuous Fourier transform of the piecewise-constant approximation of an image instead of DFT. For this, an algorithm for *topological filtration* of the spectrum of a digital image is presented. The complexity of the algorithm is the same as for fast Fourier transform [1].

This research is funded by RFBR, grants 13-07-00327 and 13-07-13112-ofi_m_RZhD.

- [1] Karkishchenko A., Mnukhin V. 2014. Topological filtration for digital images recognition and symmetry analysis. *Machine Learning and Data Analysis* 1(8):966–987. <http://jmla.org/papers/doc/2014/no8/Mnukhin2014Topological.pdf>.

Комплексирование информации с использованием N-грамм при распознавании символьных последовательностей на видеопоследовательностях

*Каратеев Сергей Львович** goga@gosniias.ru
Костромов Никита Алексеевич nikita-kostromov@yandex.ru
Бекетова Ирина Валентиновна irus@gosniias.ru
Визильтер Юрий Валентинович viz@gosniias.ru
Москва, ФГУП ГосНИИ авиационных систем

Распознавание изображений номеров или идентификационных кодов, требует высокой точности при считывании и распознавании каждого символа. Изображения подобных символьных последовательностей не несёт избыточной информации. Изменение любого символа ведёт к появлению нового содержания, которое, в принципе имеет смысл. строка такого типа. Коррекция неизбежных ошибок распознавания достигается за счёт увеличения объёма входной информации. Это позволяет выявить внутреннюю структуру и оценить вероятностные характеристики. Алгоритм определения исходной информации основывается на генетическом алгоритме, осуществляющем управляемую эволюцию предварительно сформированных правдоподобных гипотез. Формирование гипотез о содержании символьной последовательности и разбиение множества входных символьных строк на массивы эквивалентных слов осуществляется с использованием метода N-грамм. Для оценки эквивалентности анализируемых слов и поиска наилучшего значения используются функционалы на основе расстояния Левенштейна. Приведены результаты работы алгоритма с реальными записями, полученными при работе системы распознавания номеров вагонов и модельными данными.

- [1] *Каратеев С.Л.* Комплексирование информации с использованием N-грамм при распознавании символьных последовательностей на видеопоследовательностях, Москва: ООО Издательский дом Спектр, 2014. — С.15–24. <http://vkit.ru/papers/doc/2014/no6/Karateev2014.pdf>.

The complex information processing based on U-gramms for error correction in recognized video sequences of symbols

*Karateev Sergey**

irus@gosniias.ru

Kostromov Nikita

nikita-kostromov@yandex.ru

Beketova Irina

irus@gosniias.ru

Vizilter Yuri

viz@gosniias.ru

Moscow, State Research Institute of Aviation Systems

For recognition of identification codes or numbers images, require high accuracy during readout and recognition of each symbol. The visual images of corresponding sequences of symbols usually does not contain redundant information. A change in any symbol leads to the appearance of the new content, which, in principle makes sense. The correction of the inevitable errors of recognition is achieved due to an increase in the volume of input information. This makes it possible to reveal internal structure and to estimate probability characteristics. The algorithm of the determination of initial information is based on the genetic algorithm, achieves the controlled evolution of the preliminarily formed plausible hypotheses. The formation of hypotheses about the content of symbolic sequence and the partition of the set of input symbolic lines into the massifs of equivalent words is accomplished with the use of a method of u - grams. For evaluating the equivalence of the analyzed words and search for the optimum value are used the functionals on the basis of the Levenshteyn distance. The results of the work of algorithm with the real records, obtained with the work of the system of the recognition of the numbers of railroad cars and model data is proposed.

- [1] *Karateev S.* The complex information processing based on U-gramms for error correction in recognized video sequences of symbols // Herald of Computer and Information Technology, Moscow: Publishing house Spectrum, 2014. — p. 15–24. <http://vkit.ru/papers/doc/2014/no6/Karateev2014.pdf>.

Использование кластерного анализа для цифровой стабилизации видеопоследовательностей

*Карпов Роман Олегович**

karpovro@gmail.com

Копылов Андрей Валериевич

And.Kopylov@gmail.com

Тула, Тульский государственный университет

Важной составляющей частью практически любой системы видеонаблюдения является подсистема стабилизации видео, задача которой заключается в компенсации нежелательного межкадрового смещения, которое может присутствовать из-за вибрации при съемке. Предлагается ввести дополнительный шаг в процесс цифровой видеостабилизации — кластеризацию в двумерном пространстве, сформированном векторами движения локальных особенностей на последовательных кадрах. Кластеризация позволяет эффективно разделить глобальное движение камеры от локального движения объектов и удалить ошибочно построенные векторы движения. Введение кластеризации позволяет снизить требования к точности и устойчивости метода построения векторов движения и в то же время не требует использования сложных и точных алгоритмов кластеризации. В отличие от сегментации или процедуры RANSAC, часто используемых для выделения глобального смещения, для оценки параметров движения можно использовать любой из полученных кластеров по выбору пользователя. Использование самого наполненного кластера, как правило, соответствует стабилизации движения фона, а использование какого-либо другого кластера — стабилизации на объекте или группе объектов, имеющих схожее направление движения. Для выделения нежелательной случайной составляющей движения используется фильтр Калмана. В итоге для компенсации случайного движения камеры к текущему кадру применяется аффинное преобразование с оцененными параметрами [1].

Работа поддержана грантом РФФИ № 12-07-92000.

- [1] *Карпов Р. О. Копылов А. В.* Использование кластерного анализа для цифровой стабилизации видеопоследовательностей // Машинное обучение и анализ данных, 2014 (в печати).

Clustering for digital video stabilization

*Karpov Roman**

karpovro@gmail.com

Kopylov Andrey

And.Kopylov@gmail.com

Tula, Tula State University

Video stabilization is one of the necessary parts of almost each object tracking or video surveillance system. The goal of video stabilization is to compensate unwanted interframe transformations that can be produced by camera shakes and jitter or can be caused by disadvantages of camera mount and so on. Here, the additional stage of a video stabilization process, namely, cluster analysis in a two-dimensional space formed by the interframe motion of local features, is introduced. Such clustering lets to separate local and global motion and remove wrong motion vectors. In the considered case, clustering allows to reduce requirements for both motion vectors construction and clustering algorithm accuracy. As opposite to the RANSAC procedure or a segmentation that are frequently used to separate the global motion, each cluster selected by user can be used to estimate the motion parameters. The choice of the most rich cluster corresponds, as a rule, to the background motion stabilization, and using other clusters corresponds to an object or same moving objects motion stabilization. Kalman filter is used to separate an undesirable random motion component. As a result, affine transform with estimated parameters is applied to the current frame to compensate the unwanted motion [1].

This research is funded by RFBR, grant 12-07-92000.

- [1] Karpov R., Kopylov A. 2014 (in press). Clustering for digital video stabilization. *Machine Learning and Data Analysis*.

Алгоритм автоматического совмещения сенсорной и синтезируемой видеоинформации для авиационной системы комбинированного видения

<i>Лебедев Максим Алексеевич*</i>	lebedev_maxim@list.ru
<i>Бондаренко Максим Андреевич</i>	max.bond@bk.ru
<i>Комаров Денис Валерьевич</i>	mrkomap@mail.ru
<i>Степаньянц Дмитрий Георгиевич</i>	robin8081@gmail.com
<i>Выголов Олег Вячеславович</i>	o.vygovlov@gosniias.ru
<i>Визильтер Юрий Валентинович</i>	viz@gosniias.ru
<i>Желтов Сергей Юрьевич</i>	zhl@gosniias.ru

Москва, Государственный научно-исследовательский институт авиационных систем

Одной из актуальных задач интеллектуализации обработки видеоинформации в авиационных системах информационной поддержки экипажа является обеспечение режима «комбинированного видения» закабинного пространства на основе совмещения двух изображений — сенсорного (от бортового датчика технического зрения) и трехмерного (3D) изображения местности, синтезируемого с использованием значений пилотажно-навигационных параметров воздушного судна, измеряемых с погрешностями.

Предложен алгоритм автоматического совмещения сенсорного и синтезированного изображений, включающий три основных этапа: обнаружение на сенсорном изображении точек взлетно-посадочной полосы (ВПП) с использованием модифицированного преобразования Хафа, внешнее ориентирование виртуальной камеры по точкам ВПП, синтез 3D изображения с учетом поправок процедуры внешнего ориентирования. Приведены примеры и результаты работы алгоритма при моделировании реальных траекторий захода на посадку [1].

Работа поддержана грантом РФФИ № 13-08-01071.

- [1] *Лебедев М. А., Бондаренко М. А., Комаров Д. В., Степаньянц Д. Г., Выголов О. В., Визильтер Ю. В., Желтов С. Ю.* Алгоритм автоматического совмещения сенсорной и синтезируемой видеоинформации для авиационной системы комбинированного видения // Вестник компьютерных и информационных технологий, 2014 (в печати).

An algorithm for real and synthetic images matching with application to aviation combined vision system

<i>Lebedev Maksim*</i>	lebedev_maxim@list.ru
<i>Bondarenko Maksim</i>	max.bond@bk.ru
<i>Komarov Denis</i>	mrkomap@mail.ru
<i>Stepaniants Dmitriy</i>	robin8081@gmail.com
<i>Vygolov Oleg</i>	o.vygolov@gosniias.ru
<i>Vizilter Yuriy</i>	viz@gosniias.ru
<i>Zheltoy Sergey</i>	zhl@gosniias.ru

Moscow, State Research Institute of Aviation Systems

A Combined Vision System (CVS) display concept is among the urgent tasks of intelligent video processing in aircraft systems, which intend to provide information support of a flight crew. It combines a real image from onboard optical sensor with a three-dimensional computer-generated synthetic image. To create the CVS image, a problem of real and synthetic images matching should be solved, since the latter is generated by relying on the information of an aircraft current position which is measured with errors.

The paper presents an algorithm for real and synthetic images matching based on runway points detection using modification of Projective Hough Transform. The exterior orientation of virtual camera is performed using the detected runway points to deliver a pixel correspondence between the real and synthetic images.

The paper contains some experimental results obtained on simulated data. It is shown that on the base glidepath, the algorithm provides CVS image with required accuracy — even in the case of significant navigation errors (up to 10°) [1].

This research is funded by RFBR, grant 13-08-01071.

- [1] Lebedev M., Bondarenko M., Komarov D., Stepaniants D., Vygolov O., Vizilter Yu., Zheltov S. 2014 (in press). An algorithm for real and synthetic images matching with application to aviation combined vision system. *Herald of Computer and Information Technologies*.

Метод встраивания информации в видеопоследовательности, стойкий к ошибкам потери синхронизации

*Митекин Виталий Анатольевич**
Федосеев Виктор Андреевич

mitekin@smr.ru
vicanfed@gmail.com

Самара, ИСОИ РАН

В работе представлен новый метод стеганографического встраивания информации в цифровые видеопоследовательности, не требующий временной синхронизации исходной видеопоследовательности и видеопоследовательности со встроенной информацией. Данная особенность метода обуславливает высокую устойчивость метода к непреднамеренным и намеренным ошибкам десинхронизации — т. е. к пропуску и замене отдельных кадров, а также к изменению порядка следования кадров.

В работе проведен ряд вычислительных экспериментов, показавших стойкость метода как к непреднамеренным ошибкам десинхронизации, так и к преднамеренным атакам, направленным на обнаружение и приближенное вычисление встроенной информации без знания ключа. Показано, что предложенный метод встраивания позволяет восстанавливать встроенную информацию при удалении до 30% кадров исходной видеопоследовательности. Кроме того, в ходе экспериментов показана устойчивость метода к наиболее типичным для видеопоследовательностей искажениям (сжатие с потерями, кадрирование). Было проведено сравнение предложенного метода с методом кодирования Дэйви и МакКея, также обеспечивающим корректное восстановление встроенной информации при наличии ошибок десинхронизации. Показано, что для ограниченных объемов встраиваемой информации (до 704 бит) предложенный метод обладает большей кодовой скоростью, т. е. обеспечивает больший объем встраиваемой информации на кадр видеопоследовательности [1].

Работа выполнена при поддержке гранта президента РФ МК-3863.2013.9.

- [1] *Митекин В. А. Федосеев В. А.* Метод встраивания информации в видеопоследовательности, стойкий к ошибкам потери синхронизации // *Компьютерная оптика*, 2014 (в печати).

A new method for high-capacity information hiding in video robust against temporal desynchronization

*Mitekin Vitaly**

mitekin@smr.ru

Fedoseev Victor

vicanfed@gmail.com

Samara, IPSI RAS

This paper presents a new method for high-capacity information hiding in digital video and also algorithms of embedding and extraction of hidden information based on this method. These algorithms do not require temporal synchronization to provide robustness against both malicious and nonmalicious frame dropping (temporal desynchronization). Also, due to randomized distribution of hidden information bits across the video frames, the proposed method provides hidden information robustness against “watermark estimation” attack aimed at detection or extraction of hidden information without knowing the embedding key. The presented experimental results demonstrate this feature of the proposed method.

Experimental investigation also has shown method’s robustness against frame dropping (hidden information can be extracted correctly after deleting up to 30% of host video frames). At the same time, the proposed method provides high robustness against common video processing operations (lossy compression using MPEG-4 and MJPEG algorithms, frame cropping).

Also, the experiments showed that the proposed algorithm allows to achieve a higher code rate compared to the redundant coding algorithm for information transmission over channels with insertions, deletions, and substitutions (Davey and MacKay) for message length up to 704 bits [1].

This research is funded by the Russian President Grant for young scientists MK-3863.203.9.

- [1] Mitekin V., Fedoseev V. 2014 (in press). A new method for high-capacity information hiding in video robust against temporal desynchronization. *Computer Optics*.

Разработка исследовательской программной платформы для параллельного объектно-ориентированного логического программирования интеллектуального видеонаблюдения

<i>Морозов Алексей Александрович</i> ^{1,2*}	morozov@cplire.ru
<i>Вайш Абишек</i> ³	abhishek@iiita.ac.in
<i>Полупанов Александр Фёдорович</i> ^{1,2}	sashap55@mail.ru
<i>Анциперов Вячеслав Евгеньевич</i> ¹	antciperov@cplire.ru
<i>Лычков Игорь Игоревич</i> ⁴	lychkovi@gmail.com
<i>Алфимцев Александр Николаевич</i> ⁴	alfim@bmstu.ru
<i>Девятков Владимир Валентинович</i> ⁴	deviatkov@iu3.bmstu.ru

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

²Москва, Московский городской психолого-педагогический университет

³Аллахабад, Индийский институт информационных технологий

⁴Москва, МГТУ им Н. Э. Баумана

Рассмотрен метод интеллектуального видеомониторинга аномального поведения людей на основе логического программирования. Метод основан на использовании логики первого порядка для описания различных случаев аномального поведения, таких как драка, оставленный предмет, праздношатание, неподвижный человек и др. Для технической поддержки исследований в области интеллектуального видеонаблюдения создана программная платформа на основе параллельного объектно-ориентированного логического языка Акторный Пролог и транслятора из Акторного Пролога в Джаву. Обсуждается проблема автоматического синтеза логических правил распознавания сложного поведения объектов. Рассмотрена проблема разработки встроенных классов Акторного Пролога для низкоуровневой обработки видео [1].

Проект поддержан РФФИ, грант 13-07-92694, а также Министерством науки и техники Индии, грант DST P-159.

- [1] *Morozov A. A., Vaish A., Polupanov A. F., Antciperov V. E., Lychkov I. I., Alfimtsev A. N., Deviatkov V. V.* Development of concurrent object-oriented logic programming system to intelligent monitoring of anomalous human activities // Conference (International) on Biomedical Electronics and Design Proceedings, 2014. doi: 10.5220/0004929400530062.

Development of concurrent object-oriented logic programming platform for the intelligent monitoring of anomalous human activities

Morozov Alexei^{1,2*}

morozov@cplire.ru

*Vaish Abhishek*³

abhishek@iita.ac.in

Polupanov Alexander^{1,2}

sashap55@mail.ru

Antsiperov Viacheslav

antciperov@cplire.ru

*Lychkov Igor*⁴

lychkovi@gmail.com

*Alfimtsev Aleksandr*⁴

alfim@bmstu.ru

*Deviatkov Vladimir*⁴

deviatkov@iu3.bmstu.ru

¹Moscow, Kotelnikov Institute of Radio Engineering and Electronics RAS

²Moscow, Moscow State University of Psychology and Education

³Allahabad, Indian Institute of Information Technology

⁴Moscow, Bauman Moscow State Technical University

The logic programming approach to the intelligent monitoring of anomalous human activity is considered. The main idea of this approach is to use first-order logic for describing abstract concepts of anomalous human activity, i. e., brawl, sudden attack, armed attack, leaving object, loitering, pickpocketing, personal theft, immobile person, etc. A research led software platform based on the Actor Prolog concurrent object-oriented logic language and a state-of-the-art Prolog-to-Java translator for examining the intelligent visual surveillance has been created. A method of logical rules creation is considered in relation to the analysis of anomalous human behavior. The problem of creation of special built-in classes of Actor Prolog for the low-level video processing is discussed [1].

This research is funded by the Russian Foundation for Basic Research, grant 13-07-92694, and Department of Science and Technology, Government of India, grant DST-RFBR P-159.

- [1] Morozov A. A., Vaish A., Polupanov A. F., Antsiperov V. E., Lychkov I. I., Alfimtsev A. N., Deviatkov V. V. 2014. Development of concurrent object-oriented logic programming system to intelligent monitoring of anomalous human activities. *Conference (International) on Biomedical Electronics and Design Proceedings*. doi: 10.5220/0004929400530062.

Алгебраические свойства операторов распознавания в моделях зрительного восприятия динамических сцен

Панов Александр Игоревич

pan@isa.ru

Москва, ИСА РАН

Рассматривается задача моделирования зрительного восприятия. Приводятся основные принципы построения нисходящих моделей, к которым относятся иерархичность, способность выдвигать гипотезы, способность распознавать как статистические, так и динамические сцены, управляемость. На основе этих принципов приводится формальное определение распознающего блока как базового математического объекта большинства моделей восприятия. Описывается алгоритм его работы в терминах теории управления. На основе алгоритма строятся два типа операторов распознавания: статические и динамические [1]. Приводится постановка классической статической и динамической задач распознавания и исследуются свойства корректности построенных операторов с использованием алгебраического подхода Ю. И. Журавлева.

Работа поддержана грантом РФФИ № 14-17-31194.

- [1] *Панов А. И.* Алгебраические свойства операторов распознавания в моделях зрительного восприятия динамических сцен // Машинное обучение и анализ данных, 2014. Т. 1. № 7. С. 863–874. <http://jmla.org/papers/doc/2014/no7/Panov2014Algebraic.pdf>.

Algebraic properties for recognition operators in modeling visual perception of dynamic scenes

Panov Aleksandr

pan@isa.ru

Moscow, ISA RAS

The article discusses the problem of modeling of visual perception and main principles of top-down models: hierarchy, the ability to propose hypotheses, the ability to recognize both statistical and dynamic scenes, and controllability. On the basis of these principles, the formal definition of a recognition unit as a basic mathematical object of most models of perception is proposed. An algorithm for this work is described in terms of control theory. On the basis of this algorithm, two types of recognition operators (static and dynamic) are constructed [1]. Classical static and dynamic formulations of the recognition problem are performed and algebraic properties of operators constructed are studied using the theory of Yu. I. Zhuravlev.

This research is funded by RFBR, grant 14-17-31194.

- [1] Panov A. 2014. Algebraic properties fo recognition operators in modeling visual perception of dynamic scenes. *Mashine Learning and Data Analysis* 1(7):863–874. <http://jmla.org/papers/doc/2014/no7/Panov2014Algebraic.pdf>.

Использование взаимной информации как целевой функции качества оценивания параметров изображений

Ташлинский Александр Григорьевич

tag@ulstu.ru

*Воронов Сергей Васильевич**

s.voronov@ulstu.ru

г. Ульяновск, ФГБОУ ВПО УлГТУ

Рассмотрено использование в псевдоградиентных процедурах оценивания параметров изображений в качестве целевой функции взаимной информации (ВИ). Основой нахождения псевдоградиента ВИ является оценка плотности распределения вероятностей (ПРВ) яркостей изображений по локальной выборке. Хорошее качество оценивания ПРВ обеспечивает метод окна Парзена. Однако он дает низкую точность нахождения псевдоградиента ВИ и требует значительного объема локальной выборки. С целью сокращения вычислительных затрат предложен способ оценки энтропии изображений, основанный на процедуре скользящего контроля и позволяющий по сравнению с известным алгоритмом ЕММА (Efficient Mixed-Model Association) сократить объем выборки до 2 раз. При этом вычислительные затраты на нахождение энтропии изображений сокращаются на 10%–12%. При использовании метода окна Парзена оценку производных яркости отсчетов по пространственным координатам приходится искать через конечные разности. Точность этих оценок, а в конечном итоге и псевдоградиента ВИ, зависит от приращений, используемых в конечных разностях. Проанализированы возможности минимизации погрешности нахождения ПРВ яркостей изображений за счет оптимизации величины приращений. Показано, что оптимальные значения определяются видом корреляционной функции изображений и отношением сигнал/шум. При псевдоградиентном оценивании параметров межкадровых геометрических деформаций изображений использование оптимизации дает почти двукратный выигрыш в быстродействии [1].

Работа поддержана грантом РФФИ № 13-01-00555-а.

- [1] *Ташлинский А. Г., Воронов С. В.* Использование взаимной информации как целевой функции качества оценивания параметров изображений // Автоматизация процессов управления, 2014. № 3.

On the use of mutual information as objective function for image parameters' estimation

Tashlinskiy Alexander

tag@ulstu.ru

*Voronov Sergey**

s.voronov@ulstu.ru

Ulyanovsk, Ulyanovsk State Technical University

The use of mutual information (MI) is considered as an objective function for stochastic gradient image parameters' estimation. One of the main steps of mutual information gradient estimation is estimating the probability density function (PDF) of image intensities using image sample. Parzen window estimation provides a good estimation quality. However, it gives low mutual information gradient estimation accuracy and requires relatively big sample. In order to reduce the computational cost, use of a rotation estimation procedure for image entropy calculation is suggested. In comparison with well-known EMMA (Efficient Mixed-Model Association) algorithm, it allows to reduce the sample used for calculation twice. This results in reducing the computational cost of entropy estimation approximately by 10%–12%. When using the Parzen window algorithm, the derivatives of samples' intensities in the spatial coordinates should be estimated using the finite differences. The accuracy of these estimates and MI gradient estimation depend on the increments used in finite differences. The possibilities for error minimization finding PDF image brightness by optimizing the values of the increments are analyzed. It is shown that the optimal values are determined by the form of the correlation function of images and signal-to-noise ratio. When using stochastic gradient descent algorithm for image registration, it results in double gain in performance [1].

This research is funded by RFBR, grant 13-01-00555-a.

- [1] Tashlinskiy A. G., Voronov S. V. 2014. On the use of mutual information as objective function for image parameters' estimation. *Automation of Control Processes* 3.

О зависимости точности морфологического алгоритма обнаружения груза на вилочном погрузчике от интенсивности аддитивного шума

*Черноусов Владимир Олегович**

v.chernousov@mail.ru

Савченко Андрей Владимирович

avsavchenko@hse.ru

Нижний Новгород, НИУ Высшая школа экономики

Исследуется задача обнаружения движущегося вилочного погрузчика на видео при наличии резких изменений освещенности при невысоком разрешении данных, в которой точность традиционного сопоставления локальных дескрипторов (SURF — speeded up robust features, SIFT — scale-invariant feature transform, FAST — features from accelerated segment test, ORB — oriented FAST and rotated BRIEF) не достаточна.

Предложен новый алгоритм, на первом этапе которого на кадре выделяются движущиеся объекты с помощью метода МНН (motion history image), после чего на передней части объекта находится потенциальная область вил и груза. На втором этапе оператором Кенни выделяются контуры, затем с помощью морфологических преобразований вычисляются элементарные геометрические признаки объекта. Показано, что подход, использующий ограничивающие прямоугольники в качестве критерия поиска пустых вилок погрузчика, позволяет на 40% повысить точность детектирования груза по сравнению с локальными дескрипторами и является устойчивым к аддитивному шуму, а обработка одного кадра происходит в среднем на 30 мс быстрее. Результаты получены на видеоданных компании ISS (Intelligence Secure Systems, <http://www.iss.ru/>) и представлены в табл. 1 [1].

Таблица 1. Сравнение эффективности алгоритмов

Алгоритм	Точность, %	Время обработки кадра, мс
Текстурный (FAST)	44,5	44,5
Морфологический	78,7	5,2

- [1] Черноусов В. О., Савченко А. В. Помехоустойчивый морфологический алгоритм обнаружения вилочного погрузчика на видео // Машинное обучение и анализ данных, 2014 (в печати).

Dependence of the moving forklift truck detection morphological algorithm accuracy on the intensity of the additive noise

*Chernousov Vladimir**

v.chernousov@mail.ru

Savchenko Andrey

avsavchenko@hse.ru

Nizhny Novgorod, National Research University Higher School of Economics

The problem of video-based detection of the moving forklift truck is explored. It is shown that the detection quality of the state-of-the-art local descriptors (SURF — speeded up robust features, SIFT — scale-invariant feature transform, FAST — features from accelerated segment test, ORB — oriented FAST and rotated BRIEF) is not satisfactory if the resolution is low and the lighting is changed dramatically.

A new algorithm which first step is the estimation of the moving objects with the updating motion history image (MHI) and obtaining of the front part of the moving object with the potential area of empty lift or cargo is suggested. The second step involves Canny contour detection and binary morphological operations in front of the moving object to estimate simple geometric features of empty forklift. The experimental study shows that the approach which uses the bounding rectangles of empty forklift contours as an object validation rule gives the 40% more accurate and additive noise resistant results comparing with the best local descriptor and the average frame processing time is 30 ms faster. The results are obtained with ISS (Intelligence Secure Systems, <http://www.iss.ru/>) videodata and are presented in Table 1.

Table 1. Algorithm efficiency comparison

Algorithm	Accuracy, %	Frame processing time, ms
Textural (FAST)	44.5	44.5
Morphological	78.7	5.2

- [1] Chernousov V., Savchenko A. 2014 (in press). A noise-resistant morphological algorithm of video-based moving forklift truck detection. *Machine Learning and Data Analysis*.

Технология получения стандартного изображения лица с использованием программно-аппаратного комплекса подготовки и контроля цифровых фотографий и 3D-модели поверхности лица

Бекетова Ирина Валентиновна * irus@gosniias.ru

Каратеев Сергей Львович goga@gosniias.ru

Ососков Михаил Владимирович mvo@gosniias.ru

Костромов Никита Алексеевич nikita-kostromov@yandex.ru

Москва, ФГУП ГосНИИ авиационных систем

Для надежного распознавания лиц в системах компьютерного зрения целесообразно использовать изображения, удовлетворяющие требованиям ГОСТ ИСО/МЭК 19794-5-2006. Геометрически нормализованное изображение человеческого лица, характеризующееся определенными размерами и местоположением глаз, а также оценка параметров изображения, лица формируется с помощью программно-аппаратного комплекса подготовки и контроля цифровых фотографий «ЭФЕС». Представлена технология получения фронтального изображения лица на основе совмещения информации от геометрически нормализованного изображения лица и обобщенной 3D-модели поверхности лица. Разработаны оригинальные алгоритмы автоматического поиска ключевых точек изображения лица и преобразования обобщенной 3D-модели поверхности лица для приведения к форме исходного лица. Восстановление стандартного фронтального изображения лица осуществляется в реальном времени и в автоматическом режиме.

- [1] *Бекетова И. В.* Технология получения стандартного изображения лица с использованием программно-аппаратного комплекса подготовки и контроля цифровых фотографий и 3D-модели поверхности лица // Вестник компьютерных и информационных технологий, Москва: ООО Издательский дом Спектр, 2014. — С. 5–14. <http://vkit.ru/papers/doc/2014/no6/Beketova2014.pdf>.

Technology standard picture person using hardware-software complex preparation and control of digital images and 3D-model of the surface faces

Beketova Irina *

irus@gosniias.ru

Karateev Sergey

goga@gosniias.ru

Ososkov Michael

mvo@gosniias.ru

Kostromov Nikita

nikita-kostromov@yandex.ru

Moscow, State Research Institute of Aviation Systems

For reliable face recognition in computer vision systems it is advisable to use the images that satisfy the requirements of GOST ISO / IEC 19794-5-2006. Geometrically normalized image of a human face, characterized by determining the size and location of the eyes, as well as estimation of the parameters of the image, the face is formed by the hardware-software complex preparation and control of digital images “Efes”. The technology of producing frontal face image based on a combination of information from the geometrically normalized face image and generalized 3D-model of the surface of the face. Original algorithms developed automatic search key points of the face image and transform generalized 3D-model of the surface to bring the person to the shape of the original face. Restore standard frontal face image is done in real time and automatically.

- [1] *Beketova I.* Technology standard picture person using hardware-software complex preparation and control of digital images and 3D-model of the surface faces // Herald of Computer and Information Technology, Moscow: Publishing house Spectrum, 2014. — p.5–14. <http://vkit.ru/papers/doc/2014/no6/Beketova2014.pdf>.

Применение процедуры сравнения скелетов бинарных изображений к задаче распознавания типов лекарственных растений

Кушнир Олеся Александровна kushnir-olesya@rambler.ru
Середин Олег Сергеевич oseredin@yandex.ru

Тула, Тульский государственный университет

Рассматривается решение практических задач классификации при помощи предложенного метода сравнения бинарных изображений на основе цепочек примитивов с учетом радиальной функции скелета [1]. Идея учета состоит в параметризации функции ширины скелетного ребра вектором интерполяционных коэффициентов Лежандра.

Приводятся результаты экспериментальных исследований по классификации базы листьев лекарственных растений, состоящей из 32 классов, а также еще двух баз листьев, часто используемых в научных исследованиях: базы, содержащей 100 видов листьев, и довольно известной базы «*Leaves from Swedish Trees*», представляющей 15 классов листьев. Оценивается качество классификации изображений для векторов коэффициентов Лежандра с числом компонент от 4 до 8 и для схемы классификации по трем и пяти ближайшим соседям. В среднем качество распознавания составляет 77%. Неудовлетворительный результат для некоторых типов объясняется вариабельностью силуэтов внутри класса и, зачастую, высоким сходством между классами листьев овальной формы.

Для детального анализа межклассовых ошибок рассматриваются несимметрические таблицы сопряженности (contingency tables). На основе таких таблиц можно сделать выводы о внутриклассовых различиях, межклассовом сходстве, хорошо отделимых классах, о ситуациях, когда один класс «поглощает» другой, и пр.

Работа поддержана грантами РФФИ № 14-07-31271-мол_а, 14-07-00527-А и 12-07-92000-ННС_а.

- [1] Кушнир О. А. Построение функции ширины скелета бинарного изображения на основе параметрического описания многочленами Лежандра // «Известия ТулГУ», серия «Естественные науки», 2014 (в печати). Вып. 2.

Applying the procedure of binary images comparison based on skeletons to the task of medicinal leaves recognition

*Kushnir Olesia**

kushnir-olesya@rambler.ru

Seredin Oleg

oseredin@yandex.ru

Russia, Tula, Tula State University

The solution to the problem of classification tasks is suggested via the proposed method for binary images comparison based on primitive chains and radial skeleton function incorporated into primitive. The interpolation of radial function width by the vector of Legendre coefficients is the main idea which can be helpful for description of the shapes with equal skeleton topology and different width [1].

The experimental study of classification on different datasets was done: the first is the medicinal leaves dataset which consists of 32 classes, the second — the dataset consisting of 100 species, the third — is well-known database “*Leaves from Swedish Trees*” that contains 15 classes of leaves. Accuracy of image classification with Legendre coefficients number from 4 till 8 and classification scheme with 3- and 5-nearest neighbors is given as a result. The average accuracy is about 77%. Inadequate result for some species is referred to high intraspecies variation within some classes and the low interspecies variation between types with ovate leaves.

For more detailed analysis of misclassification, it makes sense to examine the nonsymmetric contingency tables for each dataset. From these tables, one can evaluate different parameters of classifier quality: for example, the interclass dissimilarity and the intraclass variance.

This research is funded by RFBR, grants 14-07-31271, 14-07-00527, and 12-07-92000.

- [1] Kushnir O., Seredin O. 2014. Parametric description of skeleton radial function by Legendre polynomials for binary images comparison. *Image and signal processing*. Eds. A. Elmoataz, O. Lezoray, F. Nouboud, D. Mammass. Lecture notes in computer science ser. Springer. 8509:520–530.

Эффективность иерархических алгоритмов классификации двумерных объектов в терминах соотношения качество–сложность

*Ланге Михаил Михайлович**
Ганебных Сергей Николаевич

lange_mm@ccas.ru
sng@ccas.ru

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

Исследуется многоклассовая метрическая модель распознавания двумерных объектов в пространстве древовидных представлений с многоуровневым разрешением. На множестве представлений построен иерархический ансамбль эталонов, отобранных по заданной мере. В ансамбле эталонов каждый класс задан последовательностью наборов мощности $m = 1, \dots, M$, в которой набор мощности m образует m -й слой бинарного дерева глубины M , а представления ансамбля эталонов на уровнях разрешения $l = 1, \dots, L$ образуют многоуровневую сеть эталонов. Предложены два решающих алгоритма по критерию ближайшего эталона на основе различных стратегий иерархического поиска в сети эталонов. Получены аналитические оценки вычислительной сложности и вычислительного выигрыша иерархических алгоритмов относительно алгоритма полного перебора эталонов как функции параметров L и M . Для иерархических и переборного алгоритмов получены экспериментальные оценки долей ошибок распознавания лиц, жестов руки и подписей. Эффективность рассмотренных решающих алгоритмов продемонстрирована сравнительными соотношениями доли ошибок от вычислительной сложности при заданном L и различных значениях M .

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

- [1] Ланге М. М., Ганебных С. Н. Иерархические структуры данных и решающие алгоритмы для классификации изображений // Машинное обучение и анализ данных, 2013. Т. 1. № 6. С. 725–733. <http://jmla.org/papers/doc/2013/no6/Lange2013ImageClassification.pdf>.

An efficiency of hierarchical algorithms for two-dimensional object classification in terms of fidelity–complexity ratio

*Lange Mikhail**

lange_mm@ccas.ru

Ganebnykh Sergey

sng@ccas.ru

Moscow, Dorodnicyn Computing Centre of RAS

A multiclass metric model of two-dimensional (2D) object recognition is investigated in a space of tree-structured multiresolution representations. In a set of the object representations, a hierarchical ensemble of template objects is constructed by a given measure. Within this ensemble, each class is described by a sequence of the template collections of cardinality $m = 1, \dots, M$ where the collection of cardinality m yields the m th layer in a binary tree of depth M . A set of the ensemble representations taken at resolution levels $l = 1, \dots, L$ generates a multilevel network of the templates. Using the nearest template criterion, two decision algorithms based on different hierarchical search strategies in the network of templates are suggested. For these algorithms, analytical estimations of a computational complexity are obtained as functions of the parameters L and M . Also, computational gains of the suggested hierarchical algorithms with respect to an exhaustive search algorithm are shown. Using the examined hierarchical and exhaustive decision algorithms, experimental estimations of recognition error rates are calculated for faces, hand gestures, and signatures. The efficiency characteristics of these algorithms are obtained in terms of comparative ratios of error rate as a function of computational complexity that are calculated for a given L and different values of M [1].

This research is funded by RFBR, grant 12-01-00920-a.

- [1] Lange M., Ganebnykh S. 2013. Hierarchical data structures and decision algorithms for efficient image classification. *Machine Learning and Data Analysis* 1(6):725–733. <http://jmla.org/papers/doc/2013/no6/Lange2013ImageClassification.pdf>.

Способы выбора наилучшего эталона радужки из последовательности

Матвеев Иван Алексеевич^{1,2}*

matveev@ccas.ru

*Новик Владимир Петрович*²

novikvp@mail.ru

¹Москва, ВЦ РАН

²Москва, ПАО ИритехКо Лтд.

В задаче идентификации по радужке предложен новый способ выбора наилучшего биометрического эталона из группы, построенной по последовательности регистрируемых изображений. Идея выбора состоит в том, что последовательность обычно включает в себя изображения с различным качеством. При этом при сравнении двух высококачественных изображений одного глаза получается малое расстояние, а если хотя бы одно из изображений имеет низкое качество, расстояние становится большим. Таким образом, изображения с высоким качеством формируют подмножество-кластер. Логично предположить, что наилучшее изображение имеет наименьшее среднее расстояние до элементов кластера (хотя расстояния до элементов, не входящих в кластер, могут быть велики).

Осуществлено сравнение со стандартным подходом, использующим признаки качества исходного изображения. Такие признаки в современных исследованиях и системах строятся для отбраковки непригодных к распознаванию изображений и могут оказаться малополезными для задачи выбора единственного лучшего изображения.

Для численных тестов использовались несколько баз изображений, находящихся в открытом доступе, общим объемом более 70 000 изображений. Выбор по расстоянию оказался лучше выбора по качеству для большинства баз. Преимуществом метода выбора по расстоянию является также то, что используется лишь функция сравнения (которая в любом случае должна присутствовать в системе распознавания) [1].

Работа поддержана грантом РФФИ № 14-01-00348.

- [1] *Матвеев И. А., Новик В. П.* Способы выбора наилучшего эталона радужки из последовательности // Машинное обучение и анализ данных, 2014. Т. 1. № 7. С. 1018–1026. <http://jmla.org/papers/doc/2014/no7/Matveev2014Iris.pdf>.

Selecting best iris template from sequence

Matveev Ivan^{1,2}★

matveev@ccas.ru

*Novik Vladimir*²

novikvp@mail.ru

¹Moscow, CC RAS

²Moscow, Iritech Inc.

A new way of selecting best template from a group of templates generated from image sequence is presented for the problem of iris identification. The idea is that the sequence typically includes images with various quality. Matching two high-quality images of one eye yields small distance, while taking at least one low-quality image makes the distance to grow. Thus, images of high quality form a subset cluster. It is logical to suppose that best image should have minimum average distance to the elements of this cluster, although distances to elements that are not in the cluster may be big.

Comparison is performed with the standard approach, which uses quality features of the source image. In the modern research and development, these features are constructed with the purpose of rejecting images unsuitable for recognition and they might be of little use for the task of selecting the best image.

Numerical tests are carried out with several public domain databases with total number of iris images over 70 000. Selection by distance happened to be superior for the majority of bases. The advantage of selection by distance is that it uses only matching function which should be present in the recognition system anyway [1].

This research is funded by RFBR, grant 14-01-00348.

- [1] Matveev I., Novik V. 2014. Selecting best iris template from sequence. *Machine Learning and Data Analysis* 1(8):1018–1026. <http://jmla.org/papers/doc/2014/no7/Matveev2014Iris.pdf>.

Меры сходства и метрики сравнения форм мозаичных изображений

*Визильтер Юрий Валентинович**

viz@gosnias.ru

Желтов Сергей Юрьевич

zhl@gosnias.ru

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

Рассмотрена задача определения процедуры морфологического сравнения мозаичных форм, согласованной с морфологическим анализом Пытьева в части сравнения изображений с формами. Исследованы четыре различных способа построения такой процедуры:

- сравнение форм на основе поиска критических точек морфологического коэффициента корреляции;
- сравнение форм на основе анализа свойств суперпозиции соответствующих им проекторов;
- сравнение форм на основе оценки раствора угла между многообразиями;
- сравнение форм на основе метрики Хаусдорфа между соответствующими многообразиями на единичной гиперсфере.

Для всех подходов независимо получены меры сходства или различия мозаичных форм. При этом меры сходства, полученные с использованием первых двух подходов (коэффициенты морфологической корреляции форм), совпали, а два других подхода привели к формированию мер отличия форм, связанных с коэффициентом морфологической корреляции форм взаимно однозначными зависимостями [1].

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

- [1] *Визильтер Ю. В., Желтов С. Ю.* Меры сходства и метрики сравнения форм мозаичных изображений // Изв. РАН. ТиСУ, 2014. № 4. С. 76. doi: 10.7868/S0002338814040167.

Similarity measures and comparison metrics for image shapes

*Vizilter Yuri**

viz@gosniias.ru

Zheltoy Sergey

zhl@gosniias.ru

Moscow, State Research Institute of Aviation Systems

The problem of defining the procedure of morphological comparison of mosaic shapes consistent with the Pytiev morphological image-to-shape comparison is discussed. Four different techniques for constructing such a procedure are studied, which are based

- on the search for critical points of the morphological correlation coefficient;
- on the analysis of properties of superposition of projectors onto shapes;
- on the estimation of the angle aperture between shapes considered as subspaces of the space of images; and
- on the Hausdorff metric between shape manifolds projected onto a unit hypersphere.

For all the techniques, measures of similarity or dissimilarity of mosaic shapes are obtained independently. The similarity measures based on the first two approaches (morphological shape correlation coefficients) turned out to be identical, while the two other techniques resulted in the dissimilarity measures that are related to the morphological correlation shape coefficient with one-to-one dependences [1].

This research is funded by RFBR, grant 14-07-00914-a.

- [1] Vizilter Yu.V., Zheltoy S.Yu. 2014. Similarity measures and comparison metrics for image shapes. *Izv. Akad. Nauk. Teoriya i sistemy upravleniya* [Theory and Control Systems] 4:76.

Морфлеты: новый класс древовидных морфологических описаний формы изображений на основе систем хаароподобных вейвлетов

<i>Горбачев Владимир Сергеевич*</i>	gvs@gosnias.ru
<i>Визильтер Юрий Валентинович</i>	viz@gosnias.ru
<i>Желтов Сергей Юрьевич</i>	zhl@gosnias.ru
<i>Рубис Алексей Юрьевич</i>	arcelt@mail.ru
<i>Воротников Андрей Валерьевич</i>	andronzord@gmail.com

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

Предложен новый класс хаароподобных вейвлетов, как систем линейных дифференциальных операторов, удовлетворяющих следующим утверждениям:

- 1) Ядра вейвлетов представляют собой кусочно-постоянные функции с нулевым средним;
- 2) Носители вейвлетов меньшего масштаба принадлежат областям постоянных значений вейвлетов большего масштаба;
- 3) Система вейвлетов образует дерево разбиений кадра.

Если такая система вейвлетов согласована с формой данного изображения (в Пытьевском смысле), то она является иерархическим описанием морфологии изображения и может использоваться для решения различных задач морфологического анализа данных. В работе введено строгое понятие хаароподобной функции. Показано, что Пытьевские морфологии центрированных изображений могут быть построены на основе хаароподобных базисов. Определены морфологии локализованных форм и соответствующие хаароподобные вейвлеты. Указан формальный способ построения иерархии локализованных форм и соответствующих деревьев хаароподобных вейвлетов. Введено понятие морфлетов как хаароподобных вейвлетов, согласованных по форме с изображением или ансамблем (классом) изображений. Описаны различные способы построения морфлетных описаний формы изображений. Работа поддержана грантом РФФИ № 14-07-00914-а.

- [1] *Горбачев В. С., Визильтер Ю. В., Желтов С. Ю., Рубис А. Ю., Воротников А. В.* Морфлеты: новый класс древовидных морфологических описаний формы изображений на основе систем хаароподобных вейвлетов // Вестник компьютерных и информационных технологий, 2014. (принято в печать)

MORPHLETS: A New Class of Tree-Structured Morphological Descriptors of Image Shape Based on Haar-like Wavelets

*Gorbatsevich Vladimir** gvs@gosniias.ru
Vizilter Yuri viz@gosniias.ru
Zheltoy Sergey zhl@gosniias.ru
Rubis Aleksey arcelt@mail.ru
Vorotnikov Andrew andronzord@gmail.com

Moscow, State Research Institute of Aviation Systems

A new class of Haar-like wavelets is proposed as a system of linear differential operators satisfying the following conditions:

- 1) Wavelet kernels are the piecewise-constant 2D functions with zero mean values;
- 2) Support (non-zero) regions of smaller scale wavelets belong to the flat (constant-valued) regions of larger scale wavelets;
- 3) System of wavelets corresponds to the tree of image frame tessellation.

If such system of wavelets corresponds to the mosaic (Pytiev) shape of some image then it will be a hierarchical morphological descriptor of this image and can be applied for solution of different problems of morphological image analysis. This paper outlines the following results. The notion of Haar-like function is formalized. It is demonstrated that Pytiev morphology of centered images (2D functions with zero mean value) can be described in terms of Haar-like basis. The morphology of localized mosaic shapes and corresponding localized Haar-like wavelet systems are described. The formal way for generation of hierarchical structure of localized mosaic shapes and corresponding tree of localized Haar-like wavelets is reported. The notion of “morphlets” as Haar-like wavelets corresponding to the mosaic shape of some image or some class of images is proposed. Some constructive schemes for generation of morphlet-based image descriptors are listed.

This research is supported by RFBR, grant 14-07-00914.

- [1] *Gorbatsevich V., Vizilter Yu, Zheltoy S., Rubis A., Vorotnikov A.* MORPHLETS: A new class of tree-structured morphological descriptors of image shape based on Haar-like wavelets // Vestnik of Computer and Information Technologies, 2014 – (in printing).

Медиальная ширина фигуры — дескриптор формы изображений

Местецкий Леонид Моисеевич

mestlm@mail.ru

Москва, Факультет вычислительной математики и кибернетики МГУ

Ширина объектов на изображениях и видеопоследовательностях может использоваться для генерации признаков при распознавании формы изображений. Известен подход к построению интегрального дескриптора ширины объекта на основе понятия *pattern spectrum* (PS), введенного Марагосом. Традиционный подход к вычислению PS для растровых изображений требует значительных затрат времени и не позволяет применить PS при обработке видеопоследовательностей и сложных изображений высокого разрешения. Работы Визильтера и Сидякина позволили существенно сократить затраты времени на вычисление PS за счет использования концепции непрерывного скелета бинарного изображения. Однако совсем исключить растровую обработку авторам не удалось. Место растровых операций открытия заняли операции дискретизации скелета и кругов медиального представления объекта. Предложена концепция построения альтернативного дескриптора формы фигуры, названного функцией медиальной ширины (ФМШ), вычисление которой осуществляется на основе непрерывного скелета и не требует применения растровых операций, что обеспечивает ускорение счета на два порядка по сравнению с PS. Функция медиальной ширины — это весовая функция точек объекта, когда вес точки полагается равным длине спицы, проходящей через точку. Спица — это радиус вписанного круга, соединяющий центр круга с точкой касания границы фигуры. Предложен алгоритм вычисления ФМШ для многоугольной фигуры и для циркулярной фигуры, получаемой скруглением углов в многоугольной фигуре. Работоспособность и эффективность предлагаемого подхода демонстрируются на примере задачи сравнения формы ладоней при биометрической идентификации личности [1].

Работа поддержана грантом РФФИ № 14-01-00716.

- [1] *Местецкий Л. М.* Медиальная ширина фигуры — дескриптор формы изображений // Машинное обучение и анализ данных, 2014 (в печати).

Medial width of figures — an image shape descriptor

Mestetskiy Leonid

mestlm@mail.ru

Moscow, MSU Faculty of Computational Mathematics and Cybernetics

Width of objects in images and video may be used for shape feature generation for image recognition. The concept of pattern spectrum (PS) introduced by P. Maragos is well suited for the role of the general descriptor of the object width. The traditional approach to the calculation of PS for raster images is based on the methods of discrete mathematical morphology. This approach is time consuming and does not allow to use PS in real-time processing for video and complex high-resolution images. Vizilter and Sidyakin proposed a method of substantially reducing time to calculate the PS through the use of the concept of continuous skeleton of a binary image. However, the authors were unable to completely exclude raster processing. Instead of the morphological opening, the raster operations for sampling of the skeleton and the medial circles came.

In this report, the concept of building an alternative shape descriptor figure called medial width function (MWF) is proposed. The method of calculation of this function based on the continuous skeleton is performed. The method does not require raster operations. This provides an acceleration of the calculations into two orders as compared with the PS. Medial width function is the weight function of object points, where the weight of the point is assumed to be the length of the spoke passing through. Spoke is the radius of the inscribed circle, connecting the circle center with a point of object boundary. An algorithm for calculating the MWF for polygonal figure and circular figure, obtained by rounding corners in a polygonal figure, is proposed. Efficiency and effectiveness of the proposed approach is demonstrated on the problem of hand-shape comparing for biometric identification [1].

The research is supported by RFBR, grant 14-01-00716.

- [1] *Mestetskiy L.* 2014 (in press). Medial width of figures — an image shape descriptor // *Machine Learning and Data Analysis*.

Косые проекторы, относительные формы и субъективные модели в морфологии изображений

Пытьев Юрий Петрович

yuri.pytyev@gmail.com

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

Математические методы и модели морфологического анализа и интерпретации изображений реальных сцен, как правило, суть математические воплощения результатов субъективного анализа цели исследования и средств ее достижения, позволяющих сформулировать морфологическую модель визуальной информации об исследуемой сцене, содержащейся в ее изображениях.

В докладе рассмотрены:

- формализм косого проецирования, определяющего относительные формы в морфологии изображений, их индексы независимости и связности [1],
- методы математического моделирования субъективных суждений исследователя об истинности относительных и абсолютных форм как моделей классов изображений, несущих правдоподобную визуальную информацию об исследуемой сцене [2];
- методы эмпирической проверки адекватности субъективной морфологической модели цели исследования [2];
- методы оптимизации решений по интерпретации результатов морфологического анализа изображений сцены [2].

Работа поддержана грантом РФФИ № 14-07-00441.

- [1] *Пытьев Ю. П.* Косые проекторы и относительные формы в морфологии изображений // Журнал вычислительной математики и математической физики, 2013. Т. 53. № 12. С. 154–176.
- [2] *Пытьев Ю. П.* Моделирование субъективных суждений модельера-исследователя о модели объекта исследования // Математическое моделирование, 2013 г., т. 25, № 4, с. 102–125.

Oblique projectors, relative forms and subjective models in image morphology

Pyt'ev Yuri

yuri.pytyev@gmail.com

Moscow, Moscow State University, Faculty of Physics

Mathematical methods and models of morphological analysis and interpretation of real scene images are, usually, mathematical implementations of the results of subjective analysis of the goal of the research and the ways of reaching it. These results allow the researcher to express the morphological model of visual information about the investigated scene contained in images of this scene.

The report covers:

- mathematical formalism of oblique projection defining:
 - relative forms in image morphology;
 - indices of independency of these relative forms; and
 - indices of interconnection of relative forms [?]
- methods of mathematical modeling of subjective judgements of the researcher about the truth of relative and absolute forms as models of image classes carrying truthful visual information about the scene under investigation [?]
- methods of empirical verification of adequacy of subjective morphological model to the goal of the research [?]and
- methods of optimizing the decisions concerning the interpretation of the results of morphological analysis of scene images [?]

This research is funded by RFBR, grant 14-07-00441.

- [1] Pyt'ev Yu. 2013. Oblique projectors and relative forms in image morphology. *Computational Mathematics and Mathematical Physics* 53(12):1916–1937. doi:10.1134/S0965542513120105.
- [2] Pyt'ev Yu. 2013. Simulation of subjective judgements made by a researcher/modeler about the model of the research object // *Mathematical Models and Computer Simulations* 5(6):538–557. doi:10.1134/S2070048213060094.

Морфологические дескрипторы формы бинарных изображений на основе эллиптических структурирующих элементов

*Сидякин Сергей Владимирович**

sersid@gosniias.ru

Визильтер Юрий Валентинович

viz@gosniias.ru

Москва, ФГУП «ГосНИИАС»

В работе Л. М. Местецкого были предложены вычислительно эффективные алгоритмы построения непрерывных скелетов плоских многоугольных фигур. Известна связь скелетных представлений с математической морфологией (ММ) Серра. В частности, морфологический спектр Марагоса с дисковым структурирующим элементом (СЭ) может быть эффективно вычислен на основе непрерывного скелетного представления и процедур rasterization «по Брезенхему». Спектр Марагоса с дисковым СЭ позволяет охарактеризовать локальную толщину частей фигуры, но при этом никак не характеризует их «направление» и «вытянутость». Поэтому существенный интерес представляет задача построения спектров с использованием СЭ, имеющих форму, отличную от дисковой, но такую, что она может быть описана аналитически и определяется конечным количеством числовых параметров, включающих фактор размера и факторы формы. В данной работе рассматривается эллиптический СЭ (ЭСЭ). Предложены понятия эллиптического скелета и эллиптических факторных спектров, а также алгоритмы их вычисления. Быстрый алгоритм построения эллиптического скелета двумерной фигуры с фиксированными факторами формы ЭСЭ основан на использовании непрерывного скелетного описания с дисковым СЭ, которое строится для эквивалентно преобразованного контура исходной фигуры. С учетом этого предложенные алгоритмы вычисления эллиптических спектров основаны на rasterization эллиптического скелетного описания. Морфологические дескрипторы для фиксированного ЭСЭ далее обобщены для произвольных ЭСЭ [1].

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

- [1] *Сидякин С. В., Визильтер Ю. В.* Морфологические дескрипторы формы бинарных изображений на основе эллиптических структурирующих элементов // Компьютерная оптика, 2014 (в печати).

Morphological shape descriptors of binary images based on elliptical structuring elements

*Sidyakin Sergey**

sersid@gosniias.ru

Vizilter Yuri

viz@gosniias.ru

Moscow, FGUP «GosNIAS»

Original computationally efficient algorithms for constructing continuous skeletons of flat polygonal figures have been proposed by L. M. Mestetskiy. At the same time, a connection between skeletal representations and Serra mathematical morphology (MM) is also known. In particular, Maragos morphological pattern spectrum with a disk structural element (SE) can be efficiently computed based on the continuous skeletal representations and Bresenham rasterizing algorithms. Maragos pattern spectrum with disk SE allows one to characterize the local thickness of figure parts, but tells nothing about the “direction” and “the degree of elongation” of these parts. Accordingly, the problem of constructing spectra of flat polygonal figures for nondisk SE is important. Upon that, SE shape should be described analytically and determined with the finite number of numerical parameters, including size and shape factors.

In this paper, elliptical SE (ellipse) is considered as such nondisk SE. The definition of elliptical skeleton and the definition of elliptical factor spectra, as well as algorithms to calculate them are proposed. Fast discrete-continuous algorithm for calculating elliptical skeletal representation with fixed SE shape factors is based on the use of continuous disk skeletal representation of equiaffine transformed polygonal contour of the original figure. After that, Bresenham rasterizing algorithms are used for fast computation of elliptical spectra with fixed SE shape factors. Morphological descriptors with fixed SE shape factors are further generalized to arbitrary elliptic SE. Application of elliptical SE allows one to deal with structural (spatial) information in addition to thickness information [1].

This research is funded by RFBR, grant 14-07-00914 A.

- [1] Sidyakin S.V., Vizilter Yu.V. 2014 (in press). Morphological shape descriptors of binary images based on elliptical structuring elements // *Computer Optics*. [In Russian.]

Автоматический подбор адекватных значений параметров регуляризации, аппроксимации и склейки скелетных графов бинарных изображений

Степанов Алексей Владимирович alexeyshagoff@gmail.com

Середин Олег Сергеевич oseredin@yandex.ru

Кушнир Олеся Александровна kushnir-olesya@rambler.ru

Тула, Тульский государственный университет

Задача выделения базовой части непрерывного скелетного графа может быть решена путем применения к нему процедур регуляризации, аппроксимации и склейки. Регуляризация скелетного графа применяется для удаления ребер, появление которых обусловлено шумами на границе фигуры. Для непрерывных скелетов используется метод регуляризации с контролируемой точностью. Суть процедуры аппроксимации состоит в уменьшении числа ребер скелета путем «спрямления» части его ветвей (ветвью называется непрерывная последовательность ребер). Склейка скелетного графа применяется для удаления избыточной информации о вершинах в предположении, что две близко расположенные смежные узловые вершины можно считать за одну. Однако обычно требуется вручную подбирать параметры данных процедур, основываясь на априорной информации о классифицируемых объектах, чтобы получить адекватное описание скелетного графа.

В данной работе рассматриваются пути решения задачи автоматического определения адекватных значений параметров регуляризации, аппроксимации и склейки. Предлагается метод, основанный на определении подходящего интервала стабилизации числа ребер скелетного графа из области определения функции зависимости этого числа от значения соответствующего параметра. Проводятся экспериментальные исследования влияния коэффициента склейки на топологию скелета.

Работа поддержана грантом РФФИ № 14-07-31271-мол_а.

- [1] *Кушнир О. А., Середин О. С., Степанов А. В.* Экспериментальное исследование параметров регуляризации и аппроксимации скелетных графов бинарных изображений // Машинное обучение и анализ данных, 2014. Т. 1. № 7. С. 817–827.

Automatic selection of the most appropriate values of pruning, approximation, and merging parameters for binary images skeletons

*Stepanov Alexey**

alexeyshagoff@gmail.com

Seredin Oleg

oseredin@yandex.ru

Kushnir Olesia

kushnir-olesya@rambler.ru

Tula, Tula State University

The problem of skeleton base subgraph extraction can be solved by pruning, approximation, and merging procedures. Pruning is the process of removing unessential skeleton ribs, which appearance is caused by noise in figure boundary. Pruning method with given accuracy is used for continuous skeletons. Approximation consists in reduction of ribs amount via replacing skeleton branches (consecutive edges) by one or several simple straight ribs. Merging is used for removing skeleton redundant nodes. The main idea is assuming that two nearby adjacent nodal vertices can be considered as one. However, usually, it is required to tune abovementioned parameters manually relying on *a priori* information about classified objects.

In this paper, the ways of automatic selection of appropriate pruning, approximation, and merging parameters values for the skeleton base subgraph extraction task are suggested. The automatic method is based on calculating the stabilization interval for all parameters. Experimental study of how the merging parameter affects the skeleton topology has been done as well.

This research is funded by RFBR, grant 14-07-31271.

- [1] Kushnir O. A., Seredin O. S., Stepanov A. V. 2014. Experimental study of regularization and approximation parameters for binary images skeletons // *Machine Learning and Data Analysis* 1(7):817–827. [In Russian.]

Обнаружение ритмов головного мозга человека на основе величины корреляции аналитических спектров электроэнцефалограммы в основных диапазонах частот

Анциперов Вячеслав Евгеньевич

antciperov@cplire.ru

Москва, Институт радиотехники и электроники
им. В. А. Котельникова РАН

Представлены результаты развития методов многомасштабного корреляционного анализа применительно к задачам обнаружения и мониторинга ритмов коры головного мозга человека. В частности, обсуждается усовершенствование техники вычисления частотно-временных представлений «конического» типа (Crooks–Kawasaki relation, СКР, являющихся частным случаем корреляционных распределений) на основе недавно разработанной автором концепции аналитических спектров [1].

В отличие от классического спектрального анализа, направленного на характеризующееся высоким разрешением оценивание индивидуальных частот, предлагаемые методы ориентированы в первую очередь на обнаружение и мониторинг колебательной активности нейронов в широких основных частотных диапазонах.

Главным результатом является синтезированное «конического» типа представление электроэнцефалограммами сигналов, имеющее вид:

$$\begin{aligned} \text{СКР}(\tau, f, t) &= 2\text{Re}\{\exp(-2\pi i\tau f)G(\tau, f, t)\}; \\ G(\tau, f, t) &= \frac{1}{\tau} \int H(f' - f|\sigma)S_{Ft}(f')S_{Pt}(f') \exp(2\pi i\tau f') df'; \\ H(f|\sigma) &= \exp\left(-\frac{(2\pi\sigma)^2}{2}f^2\right). \end{aligned}$$

Это представление допускает прозрачную интерпретацию результатов и быструю численную реализацию на основе быстрых алгоритмов преобразования Фурье для вычислений входящих под интеграл аналитических спектров и самого интеграла.

Работа поддержана грантами РФФИ № 12-07-00104 и № 14-07-00496 А.

- [1] *Анциперов В. Е.* Обнаружение ритмов головного мозга человека на основе корреляции аналитических спектров ЭЭГ в основных диапазонах частот. // Электр. журнал радиоэлектроники, 2014. № 5. С. 1–11. <http://jre.cplire.ru/jre/may14/13/text.pdf>.

Cortical rhythms detection based on the values of electroencephalography analytic spectra correlation in specific frequency ranges

Antsiperov Viacheslav

antsiperov@cplire.ru

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

Results of multiscale correlational analysis in application to electroencephalography cortical rhythms detection improvements are presented and discussed in detail. In particular, the importance of frequency-temporal cone-shaped representations calculation techniques (which is the particular case of the correlation distributions) based on the recently introduced concept of the analytical spectra is in the center of discussion.

Unlike classical spectral analysis, aimed at high-resolution assessment of individual frequencies, the proposed methods are focused primarily on detection and monitoring of oscillatory neurons activity in the wide main frequency ranges.

The most important result of the investigation is the synthesized cone-shaped representation, which has a form:

$$\begin{aligned} \text{CKR}(\tau, f, t) &= 2\text{Re}\{\exp(-2\pi i \tau f)G(\tau, f, t)\}; \\ G(\tau, f, t) &= \frac{1}{\tau} \int H(f' - f|\sigma)S_{Ft}(f')S_{Pt}(f') \exp(2\pi i \tau f') df'; \\ H(f|\sigma) &= \exp\left(-\frac{(2\pi\sigma)^2}{2}f^2\right). \end{aligned}$$

This representation assumes transparent interpretation of results and allows fast numerical realization on the basis of fast Fourier transform algorithms for calculating as included under the integral analytical spectra as the integral itself.

This research is funded by RFBR, grants 12-07-00104 and 14-07-00496 A.

- [1] Antsiperov V. 2014. Cortical rhythms detection based on EEG analytic spectra correlations in specific frequency ranges. *J. Radio Electronics* 5:1–11. <http://jre.cplire.ru/jre/may14/13/text.pdf>.

Методы визуализации многомерных медицинских данных на основе обобщенных графических образов

<i>Берестнева Ольга Григорьевна*</i>	ogb6@yandex.ru
<i>Воловоденко Виталий Алексеевич</i>	pm08@sibmail.com
<i>Гергет Ольга Михайловна</i>	olgagerget@mail.ru
<i>Шаропин Константин Александрович</i>	kashar@mail.ru
<i>Осадчая Ирина Анатольевна</i>	irishka@mail.ru

Томск, Томский политехнический университет

Целью работы являлось стремление визуализировать образы, представленные численными данными, и показать возможность выявления отношений между ними на основе наблюдений за совокупностью полученных образов. В статье рассмотрен подход, который позволяет провести визуализацию основных линейных конструкций: отрезок, ломаная линия, симплекс в многомерных пространствах. Спектральное представление визуального образа является более «тонким» инструментом, подчеркивающим различия или сходства образов, чем традиционные методы, характеризующие эти свойства на уровне числовых параметров.

Представлены результаты применения данного подхода для двух задач: (1) анализ динамики физиологического состояния беременных женщин; (2) анализ воздействия аудиовизуальной стимуляции на пациентов с различными формами бронхиальной астмы.

Применение предложенного подхода приводит к значительному упрощению процедуры визуализации и способствует выявлению скрытых связей в очевидной форме. Это позволяет не только использовать свойства метода, но и способствует учету опыта группы экспертов, рассматривающих конкретную проблему. Последнее обстоятельство повышает актуальность предложенного подхода [1].

Работа поддержана грантом РФФИ №14-07-00675

- [1] *Berestneva O. G., Volovodenko V. A., Gerget O. M., Sharopin K. A., Osadchaya I. A.* Multidimensional medical data visualization methods based on generalized graphic images // World Appl. Sci. J. №24. С. 18–23. [http://www.idosi.org/wasj/wasj\(ITMIES\)13/36.pdf](http://www.idosi.org/wasj/wasj(ITMIES)13/36.pdf).

Multidimensional medical data visualization methods based on generalized graphic images

*Berestneva Olga**

Volovodenko Vitaly

Gerget Olga

Sharopin Konstantin

Osadhaya Irina

ogb6@yandex.ru

pm08@sibmail.com

olgagerget@mail.ru

kashar@mail.ru

irishka@mail.ru

Tomsk, Tomsk Polytechnic University

The purpose of work was the aspiration to visualize the images presented by numerical data and to show the ability to identify relationships between them on the basis of observation of the totality of the obtained images. The authors suggest an approach that allows to perform visualization of the major linear structures: segment, polyline, and simplex in multidimensional spaces. The spectral representation of the visual image is more “delicate” tool which highlights the differences and similarities of images than conventional methods of characterizing these properties at the level of the numerical parameters. The results of applying this approach to the problems of practical medicine were proposed: (*i*) analysis of the pregnant women physiological state dynamics; and (*ii*) identifying hidden patterns in the structure of the physiological parameters for patients with various forms of asthma. The use of this method, in the authors’ opinion, leads to a significant simplification of the imaging procedure and helps to identify hidden relationships in an obvious way. This allows one not only use the method properties, but also contributes to the integration experience of the expert group considering a specific problem. The latter fact increases the relevance of the proposed approach [1].

This research is funded by RFBR, grant 14-07-00675.

- [1] Berestneva O. G., Volovodenko V. A., Gerget O. M., Sharopin K. A., Osadhaya I. A. 2013. Multidimensional medical data visualization methods based on generalized graphic images. *World Appl. Sci. J.* 24:18–23. [http://www.idosi.org/wasj/wasj\(ITMIES\)13/36.pdf](http://www.idosi.org/wasj/wasj(ITMIES)13/36.pdf).

Стационарные режимы в модели Хенинга и ее модификациях

Макаров Валерий Леонидович

makarov@cemi.rssi.ru

*Бекларян Левон Андреевич**

beklar@cemi.rssi.ru

Белоусов Фёдор Анатольевич

belousovfedor@gmail.com

Москва, ЦЭМИ РАН

В работе изучается динамика развития популяций в зависимости от типов взаимодействия между особями. Также изучаются условия, при которых наблюдается эффект вымирания популяций. Данная тема важна, так как подобные явления наблюдаются как в дикой природе, так и в истории человеческой цивилизации.

Первая часть доклада посвящена модели Хенинга и ее модификациям. Изучается влияние внутривидовой агрессии, в частности мести. Во второй части эти же вопросы исследуются для модели, сильно отличающейся от модели Хенинга и в большей степени похожей на хорошо известную «Сахарную модель». Главное отличие второй модели от первой — эндогенность правил взаимодействия между особями в популяции, которые определяются на основе состояния агентов и внешней среды, а не случайным образом, как в модели Хенинга. Другими словами, взаимодействие, основанное на реакции типа «инстинкт», заменяется на взаимодействие с использованием элементов «этики».

Общий вывод работы достаточно интуитивен — чем больше негативного во взаимоотношениях между особями, тем менее жизнеспособна популяция в целом. И наоборот: чем больше благотворного влияния особи оказывают друг на друга, тем более живучей оказывается популяция. В данной работе живучесть популяции исследуется не только в зависимости от позитивных и негативных факторов, но также в зависимости от характеристик как самой популяции, так и окружающей среды [1].

Работа поддержана грантом РФФИ № 12-01-00768.

- [1] *Бекларян Л. А.* Стационарные режимы в модели Хенинга и ее модификациях // Машинное обучение и анализ данных, 2014 (в печати).

Steady-state regimens in Henning model and its modifications

Makarov Valeriy

makarov@cemi.rssi.ru

*Beklaryan Levon**

beklar@cemi.rssi.ru

Belousov Fedor

belousovfedor@gmail.com

Moscow, CEMI RAS

This paper is dedicated to exploration how population dynamics depends on different types of interaction between agents. Conditions leading to death of populations are also studied. This subject is important since such phenomena can be observed in both unexplored wilderness and human civilization.

The first part of the report is dedicated to Henning model and its modifications. Aggressive behavior and, particularly, revenge are studied more attentively. In the second part, the same questions are considered by using another model which differs significantly from Henning model and to a greater extent, it is similar to well-known Sugar model. The main distinction of the second model from the first one is endogeneity of behavior rules between agents that are determined based on conditions of agents and environment, instead of determining at random, as in Henning model. In other words, interaction based on reaction of “instinct” type is replaced by using elements of “ethics.”

The general conclusion is intuitive enough — the more negative interaction between agents, the less viable population. And *vice versa*: the more positive interaction between agents, the more viable population. In this paper, population vitality is studied not only as a function of positive and negative factors, but also as a function of characteristics of the population and the environment [1].

This research is funded by RFBR, grant 12-01-00768.

- [1] Beklaryan L. 2014 (in press). Steady-state regimens in Henning model and its modifications. *Machine Learning and Data Mining*.

Облачный ресурс MathBrain для анализа данных энцефалографии

*Оплачко Екатерина Сергеевна**

oplachkoe@gmail.com

Рыкунов Станислав Дмитриевич

rykunov@psn.ru

Устинин Михаил Николаевич

ustinin@impb.psn.ru

Пушино, Институт математических проблем биологии РАН

Система MathBrain представляет собой интернет-ресурс для анализа данных энцефалографии и предоставляет пользователю инструмент для построения точного преобразования Фурье и восстановления временных рядов по заранее рассчитанному спектру. В программный комплекс входит набор скриптов Python, которые обрабатывают введенные пользователем данные на стороне сервера, используя его аппаратные мощности. Указанные скрипты используют библиотеки SAGA и менеджера ресурсов SLURM, библиотеки numpy, scipy и numexpr. Использование данного программного обеспечения обеспечивает легкую масштабируемость системы и балансировку нагрузки между пользователями. Клиентская часть реализована на языке программирования PHP в связке с языком разметки HTML, с использованием элементарных скриптов JavaScript. Веб-интерфейс позволяет пользователю загружать входные данные, задавать параметры расчета, скачивать результаты анализа и визуализировать спектры в виде графиков. Как входные данные, так и результаты расчетов сохраняются в базе данных MySQL, что позволяет вести историю запросов и накапливать экспериментальную базу. Ресурс MathBrain является многопользовательским и позволяет исследователям не задумываться над требуемыми для анализа экспериментальных данных аппаратными мощностями [1].

Работа была частично поддержана Американским фондом гражданских исследований и развития (грант RUB-7095-MO-13) и Российским фондом фундаментальных исследований (проекты РФФИ 13-07-00162, 13-07-12183, 14-07-00636 и 14-07-31309).

- [1] *Оплачко Е. С., Рыкунов С. Д., Устинин М. Н.* Комплексы прикладных программ и облачные сервисы для анализа данных энцефалографии // Математическая биология и биоинформатика, 2014. Т. 9. Вып. 2.

Cloud service MathBrain to analyze encephalography data

*Oplachko Ekaterina**

oplachkoe@gmail.com

Rykunov Stanislav

rykunov@psn.ru

Ustinin Mikhail

ustinin@impb.psn.ru

Pushchino, Institute of Mathematical Problems of Biology RAS

The MathBrain project is an online resource for encephalography data analysis which works as Software as a Service (SaaS). It performs precise Fourier transform and reconstructs time series using the calculated spectrum. The software complex includes a set of Python scripts which provide a general computational mechanism and handle users data on the server side, using its hardware capacity. These scripts use SAGA libraries and resource manager SLURM, libraries numpy, scipy, and numexpr. The use of this set of software provides easy scalability and load balancing features. The client part is implemented on the PHP programming language with HTML, using JavaScript. Web-based interface allows user to upload the experimental data with set of parameters for calculation and to download results of the analysis and visualization. All data, both input and output, are stored in the database MySQL, which allows to save history of work and to accumulate experimental database. The files in .mat format are located on the hardware resources of the system and in the database, one can see only links to them. Due to authentication feature and sessions using, MathBrain is a multiuser system, allowing researchers to use hardware facilities of the system and enjoy all the benefits of cloud computing [1].

The study was partly supported by the American Foundation of Civil Research and Development (CRDF project RUB-7095-MO-13) and the Russian Foundation for Basic Research (RFBR projects 13-07-00162, 13-07-12183, 14-07-00636, and 14-07-31309).

- [1] *Oplachko E. S., Rykunov S. D., Ustinin M. N.* 2014. Software complexes and cloud services for the encephalography data analysis. *Mathematical Biology and Bioinformatics*. 9(2).

Спектральный анализ высокочастотной области данных магнитной энцефалографии

*Панкратова Наталья Михайловна**

pnm@impb.ru

Устинин Михаил Николаевич

ustinin@impb.ru

Пушино, Институт математических проблем биологии РАН

На записях спонтанной активности головного мозга при патологии тиннитус обнаружены переключения на кратковременный высокоамплитудный режим. При этом происходит уменьшение корреляционной размерности отдельно взятого фрагмента одного временного ряда в 2 раза. Обнаруженная активность может быть отнесена к разряду пароксизмальной. Анализ собственных функций Карунена–Лоэва при наличии таких кратковременных промежутков показал, что среди первых мод присутствует паттерн, имеющий сходство с аудиторным откликом. В экспериментах с указанием той же патологии, но при отсутствии переключения между режимами, такого паттерна нет. Спектры Фурье экспериментов с переключениями демонстрируют увеличение спектральной мощности гамма диапазона. Найденная в части экспериментов высокочастотная активность имеет локализацию в слуховой области. Пароксизмальную активность с такой локализацией находят при височной эпилепсии, при которой могут наблюдаться слуховые галлюцинации. В этом случае тиннитус является лишь симптомом болезни. Экспериментальные данные, не содержащие переключений на высокоамплитудный режим, не отличаются от контрольных экспериментов по спектральной мощности в области высоких частот.

Результаты свидетельствуют о диагностической значимости обработки и анализа высокочастотной области экспериментальных данных магнито- и электроэнцефалографии, которая до сих пор остается плохо изученной.

Работа была частично поддержана проектами РФФИ № 13-07-00162, № 13-07-12183, № 14-07-00636 и № 14-07-00924.

- [1] *Панкратова Н. М., Устинин М. Н., Линас Р. Р.* Обнаружение патологической активности головного мозга по данным магнитной энцефалографии // Математическая биология и биоинформатика, 2013. Т. 8. № 2. С. 679–690. http://www.matbio.org/2013/Pankratova_8_679.pdf.

Spectral analysis of the high frequency region of magnetic encephalography data

*Pakratova Natalia**

pnm@impb.ru

Ustinin Mikhail

ustinin@impb.ru

Puschino, IMPB RAS

On recordings of spontaneous brain activity in the pathology of tinnitus, the switching to short-term high-amplitude mode is found. Thus, there is the twice decrease in the correlation dimension of a single fragment at time series. The observed activity can be classified as a paroxysmal. Analysis of the eigenfunctions of Karhunen–Loeve in presence of such short intervals showed that among the first modes, there is the pattern that resembles the auditory response. The experiments specifying the same pathology but without switching between modes have no such pattern. Fourier spectra of the switching experiments show an increase in the spectral power of gamma band. The high-frequency activity found in some experiments is localized in the auditory region of the cortex. Paroxysmal activity of such localization is found in temporal lobe epilepsy which can be observed with auditory hallucinations. In this case, the tinnitus is only the symptom of the disease. Experimental data without switching do not differ from the control experiments in the high-frequency band.

The results indicate the diagnostic importance of processing and analysis of high-frequency region of the experimental data of magneto- and electroencephalography which is still poorly understood [1].

This research is partially funded by RFBR, grants 13-07-00162, 13-07-12183, 14-07-00636, and 14-07-00924.

- [1] Pankratova N., Ustinin M., Llinas R. 2013. Detecting pathological brain activity according magnetic encephalography. *Mathematical Biology and Bioinformatics* 8(2):679–690. http://www.matbio.org/2013/Pankratova_8_679.pdf.

Выделение и локализация источников аудиторного отклика коры головного мозга по сильно зашумленным данным магнито- и электроэнцефалографии

*Рыкунов Станислав Дмитриевич**

rykunov@psn.ru

Сычев Вячеслав Викторович

sychyov@yahoo.com

Устинин Михаил Николаевич

ustinin@impb.psn.ru

Пушино, Институт математических проблем биологии РАН

Метод анализа многоканальных сигналов на основе точного преобразования Фурье был применен для исследования различных данных магнитной энцефалографии, а также данных электроэнцефалографии. Один и тот же эксперимент с тем же субъектом был проведен на двух различных типах магнитометров и на многоканальном электроэнцефалографе. Все экспериментальные наборы временных рядов были проанализированы с использованием информационной технологии, основанной на преобразовании Фурье временных рядов большой длительности и на подстройке сетки частот под известную частоту стимула. Это позволило построить систему базисных функций, оптимальную для данного эксперимента. Паттерны магнитного поля, построенные на гармониках частоты стимула, максимально очищены от шумов и дают возможность решать обратную задачу по реконструкции пространственной структуры источников поля. Было найдено, что при использовании предложенного метода анализа данных возможно проведение экспериментов по энцефалографии в условиях сильных внешних шумов [1].

Работа была частично поддержана Американским фондом гражданских исследований и развития (гранты CRDF RB1-2027 и RUB-7095-MO-13) и Российским фондом фундаментальных исследований (проекты РФФИ 13-07-00162, 13-07-12183, 14-07-00636 и 14-07-31309).

- [1] *Коршаков А. В., Поликарпов М. А., Устинин М. Н., Сычев В. В., Рыкунов С. Д., Наурзаков С. П., Гребенкин А. П., Панченко В. Я.* 2014. Регистрация и анализ точных частотных ЭЭГ/МЭГ откликов аудиторной коры в ответ на монауральную стимуляцию звуком с фиксированными частотными составляющими // Математическая биология и биоинформатика, 2014. Т. 9. Вып. 1.

Extraction and localization of the sources of the brain auditory cortical response from the noisy data of magneto- and electroencephalography

*Rykunov Stanislav**

rykunov@psn.ru

Sychyov Vyacheslav

sychyov@yahoo.com

Ustinin Mikhail

ustinin@impb.psn.ru

Pushchino, Institute of Mathematical Problems of Biology RAS

The multichannel signal analysis method based on precise Fourier transform was applied to study the magnetic encephalography data, obtained in different experimental conditions, and the electric encephalography data. The similar experiment with the same subject was performed using two different types of magnetometers and the multichannel electric encephalograph. All experimental data sets were analyzed using the same information technology, based on the Fourier transform of long time series and on the tuning of the frequency grid to the known stimulus frequency. This approach made it possible to build the system of the basic functions, optimal for the particular experiment, and to maximize the signal to noise ratio. The magnetic field patterns, calculated on the harmonics of stimulus frequency, are maximally cleaned from noises and make it possible to solve the inverse problem to reconstruct the spatial structure of the field sources. It was found that the usage of the proposed method of data analysis makes it possible to conduct encephalography experiments in case of strong external noises [1].

The study was partly supported by the American Foundation of Civil Research and Development (CRDF projects RB1-2027 and RUB-7095-MO-13) and the Russian Foundation for Basic Research (RFBR projects 13-07-00162, 13-07-12183, 14-07-00636, and 14-07-31309).

- [1] Korshakov A. V., Polikarpov M. A., Ustinin M. N., Sychyov V. V., Rykunov S. D., Naurzakov S. P., Grebenkin A. P., Panchenko V. Ya. 2014. Registration and analysis of precise frequency EEG/MEG auditory responses to monaural responses to stimulation by the sound with fixed frequency components. *Mathematical Biology and Bioinformatics* 9(1).

Совместный анализ сигналов электроэнцефалограмм, электромиограмм и механического тремора при болезни Паркинсона в ранней стадии

<i>Сушкова Ольга Сергеевна</i> ^{1*}	<i>o.sushkova@mail.ru</i>
<i>Королев Михаил Сергеевич</i> ¹	<i>korolevclub@mail.ru</i>
<i>Габова Александра Васильевна</i> ²	<i>agabova@yandex.ru</i>
<i>Иллариошкин Сергей Николаевич</i> ³	<i>snillario@gmail.com</i>
<i>Карабанов Алексей Вячеславович</i> ³	<i>doctor.karabanov@mail.ru</i>
<i>Обухов Константин Юрьевич</i> ⁴	<i>ko.arsenalfc@gmail.com</i>
<i>Обухов Юрий Владимирович</i> ¹	<i>yuvobukhov@mail.ru</i>

¹Москва, Институт радиотехники и электроники РАН

²Москва, Институт высшей нервной деятельности и нейрофизиологии РАН

³Москва, Научный центр неврологии РАН

⁴Москва, Московский физико-технический институт

С помощью вейвлет преобразования и его дальнейшего количественного анализа подтвержден ряд фактов, характеризующих особенности электроэнцефалограмм (ЭЭГ) болезни Паркинсона, и обнаружен ряд специфических особенностей частотно-временной организации ЭЭГ 1-й стадии заболевания. Гистограммы экстремумов вейвлет преобразований ЭЭГ более четко по сравнению со спектрами Фурье подчеркивают признаки болезни Паркинсона и позволяют изучать их динамику. Полученные данные намечают дальнейшие пути поиска специфических маркеров ЭЭГ самых ранних, в том числе доклинических, стадий заболевания.

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

- [1] *Сушкова О. С. Королев М. С. Габова А. В. Иллариошкин С. Н. Карабанов А. В. Обухов К. Ю. Обухов Ю. В.* Совместный анализ сигналов электроэнцефалограмм, электромиограмм и механического тремора при болезни Паркинсона в ранней стадии // Журнал радиоэлектроники, 2014. №5. <http://jre.cplire.ru/jre/may14/12/text.pdf>.

Joint analysis of the signals of electroencephalograms, electromyograms, and mechanical tremor of the Parkinson's disease in the early stage

Sushkova Olga^{1*} o.sushkova@mail.ru
*Korolev Mikhail*¹ korolevclub@mail.ru
*Gabova Alexandra*² agabova@yandex.ru
*Illarioshkin Sergey*³ snillario@gmail.com
*Karabanov Alexey*³ doctor.karabanov@mail.ru
*Obukhov Konstantin*⁴ ko.arsenalfc@gmail.com
*Obukhov Yuri*¹ yuvobukhov@mail.ru

¹Moscow, Kotel'nikov Institute of Radioengineering and Electronics of RAS

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

³Moscow, Scientific Center of Neurology of RAS

⁴Moscow, Moscow Institute of Physics and Technology

By means of the wavelet transform and its further quantitative analysis, a number of facts that characterize the electroencephalogram (EEG) features of Parkinson's disease are confirmed and a number of specific features of the time-frequency organization of EEG of the 1st stage of the disease are discovered. The extrema histograms of EEG wavelet transforms more clearly, compared with the Fourier spectra, underline symptoms of Parkinson's disease and allow to study their dynamics. The obtained data outlined further search ways of the specific EEG markers of the earliest, including preclinical, stages of the disease.

This research is funded by RFBR, grant 12-02-00611-a, and by the Program of the Presidium RAS "Fundamental sciences — for medicine."

- [1] Sushkova O., Korolev M., Gabova A., Illarioshkin S., Karabanov A., Obukhov K., Obukhov Y. 2014. Joint analysis of the signals of electroencephalograms, electromyograms, and mechanical tremor of the Parkinson's disease in early stage. *J. RadioElectronics* 5. <http://jre.cplire.ru/jre/may14/12/text.pdf>.

Информационный анализ электрокардиосигналов для диагностики многих заболеваний внутренних органов по одной электрокардиограмме

*Успенский Вячеслав Максимилианович*⁴ medddik@yandex.ru
Воронцов Константин Вячеславович^{1,2,3*} voron@forecsys.ru
*Цельх Влада Руслановна*¹ celyh@inbox.ru
*Бунаков Василий Андреевич*¹ va.bunakov@gmail.com
*Ефимова Ирина Валерьевна*¹ efimova@phystech.edu
*Полежаев Валентин Александрович*³ walter2kf@gmail.com

Москва, ¹МФТИ, ²ВЦ РАН, ³ВМК МГУ, ⁴Медицинский учебно-научный клинический центр им. П. В. Мандрыка

Технология информационного анализа электрокардиосигналов основана на наблюдении, что вариации амплитуд и интервалов кардиоциклов несут значимую информацию о состоянии не только сердца, но и всего организма. Технология включает следующие этапы анализа ЭКГ-сигнала: (1) вычисление знаков приращений интервалов и амплитуд в последовательности кардиоциклов; (2) формирование кодограммы — последовательности символов 6-буквенного алфавита, либо нечёткой кодограммы — последовательности вероятностей символов; (3) формирование числового вектора частот триграмм; (4) построение диагностических правил с использованием методов отбора признаков, логических, метрических или линейных алгоритмов классификации. Обучающая выборка включает около 20 тысяч электрокардиограмм, накопленных за 15 лет эксплуатации диагностических систем «Поток» и «Скринфакс», и охватывает более 30 заболеваний внутренних органов. В докладе приводятся статистические обоснования технологии информационного анализа для диагностики многих заболеваний по одной электрокардиограмме. Эксперименты подтверждают удивительно высокие уровни чувствительности и специфичности (от 80% до 96% в зависимости от заболевания), зачастую превышающие точность специальных методов диагностики.

Гранты РФФИ: 14-07-00847, 14-07-00908, 14-07-31163.

- [1] *Успенский В. М., Воронцов К. В., Цельх В. Р.* Статистические обоснования информационного анализа электрокардиосигналов для диагностики многих заболеваний // Математическая биология и биоинформатика, 2014 (в печати).

Informational analysis of electro cardio signals for multidisease diagnostic system

*Uspenskiy Vyacheslav*⁴ medddik@yandex.ru
Vorontsov Konstantin^{1,2,3*} voron@forecsys.ru
*Tselykh Vlada*¹ celyh@inbox.ru
*Bunakov Vasiliy*¹ va.bunakov@gmail.com
*Efimova Irina*¹ efimova@phystech.edu
*Polezhaev Valentin*³ valentin.polezhaev@gmail.com

Moscow, ¹MIPT, ²CC RAS, ³Moscow State University, ⁴P. V. Mandryk Medical Educational and Scientific Clinical Center

The informational analysis of electro cardio signals is based on an observation that variations of the amplitudes and time intervals of QRS-ventricular complexes of the electrocardiogram contain a meaningful information about many internal diseases, not only heart disease.

We propose the ECG signal processing technique, which includes following steps: (1) measuring signs of the increments of the amplitudes and time intervals in subsequent cardiac complexes; (2) transforming an electrocardiogram into a codogram — an encoded text string in the 6-letter alphabet, or into a fuzzy codogram — a sequence of probability distributions over the 6-letter alphabet; (3) transforming a codogram into a 216-dimensional numerical vector of trigram frequencies; (4) learning a diagnostic rule using machine learning techniques.

We use permutational tests for feature selection and then we learn linear, rule-based, and distance-based classifiers in order to choose a best diagnostic rule for each disease.

Training sample contains more than 20 000 ECGs for more than 30 diseases, accumulated over 15 years from clinical diagnostic systems “Potok” and “Skrinfaks”.

Experiments confirm the surprisingly high levels of sensitivity and specificity (from 80% to 96% depending on the disease), often exceeding the accuracy of specialized diagnostic methods.

Thanks to RFBR grants: 14-07-00847, 14-07-00908, 14-07-31163.

- [1] *Uspenskiy V. M., Vorontsov K. V., Tselykh V. R.* Statistical verification of electro cardio signals informational analysis for multidisease diagnostic system // *Mathematical Biology and Bioinformatics*, 2014 (to appear).

Функциональная томография мозга человека на основе анализа энцефалограмм в пространстве частота–паттерн

Устинин Михаил Николаевич

ustinin@impb.psn.ru

Пушино, Институт математических проблем биологии РАН

Предложен новый метод анализа многоканальных временных рядов. Метод основан на точном преобразовании Фурье и анализе когерентности восстановленных временных рядов для всех частотных компонент. При анализе магнитных энцефалограмм спонтанной и вызванной активности головного мозга человека было найдено, что точный многоканальный спектр Фурье содержит большое количество гармоник с высокой когерентностью. Для таких гармоник обратная задача может быть решена с высокой точностью, что определяет перспективность метода в задачах изучения и отображения активности мозга. Предложенный метод позволяет реконструировать структуру и временную зависимость как спонтанной, так и вызванной активности мозга с миллисекундной точностью [1].

Работа была частично поддержана Американским фондом гражданских исследований и развития (гранты CRDF RB1-2027 и RUB-7095-MO-13) и Российским фондом фундаментальных исследований (проекты РФФИ 13-07-00162, 11-07-00577, 13-07-12183 и 14-07-00636).

- [1] *Llinás R. R., Ustinin M. N.* Frequency–pattern functional tomography of magnetoencephalography data allows new approach to the study of human brain organization // *Front. Neural Circuits*, 2014. Vol. 8. P. 43. doi:10.3389/fncir.2014.00043 <http://journal.frontiersin.org/Journal/10.3389/fncir.2014.00043/pdf>.

Functional tomography of the human brain based on frequency-pattern analysis of the encephalography data

Ustinin Mikhail

ustinin@impb.psn.ru

Pushchino, Institute of Mathematical Problems of Biology RAS

A method based on a set of new theorems for the analysis of multi-channel time series is described based on precise Fourier transform and coherence analysis of the restored signals from a detailed set of frequency components. Magnetic field recordings of spontaneous and evoked activity by means of magnetic encephalography demonstrated that multichannel precise Fourier spectrum contains a very large set of harmonics with high coherence. The inverse problem can be solved with great precision based on coherent harmonics; so, the technique is a promising platform of general analysis in brain imaging. The analysis method makes it possible to reconstruct sites and timing of electrical activity generated by both spontaneous and evoked brain function at different depths in the brain in the millisecond time range [1].

The study was partly supported by the American Foundation of Civil Research and Development (CRDF projects RB1-2027 and RUB-7095- MO-13) and the Russian Foundation for Basic Research (RFBR projects 13-07-00162, 11-07-00577, 13-07-12183, and 14-07-00636).

- [1] Llinás R. R., Ustinin M. N. 2014. Frequency-pattern functional tomography of magnetoencephalography data allows new approach to the study of human brain organization. *Front. Neural Circuits* 8:43. doi:10.3389/fncir.2014.00043 <http://journal.frontiersin.org/Journal/10.3389/fncir.2014.00043/pdf>.

Кластер-анализ пространственных контактов аминокислотных остатков белков с нуклеотидами ДНК

*Кузнецов Евгений Николаевич*¹

ken@ipu.ru

Анашкина Анастасия Андреевна^{2*}

nastya@eimb.ru

*Есипова Наталья Георгиевна*²

nge@eimb.ru

*Туманян Владимир Гаевич*²

tuman@eimb.ru

¹Москва, Институт проблем управления им. В. А. Трапезникова РАН

²Москва, Институт молекулярной биологии им. В. А. Энгельгардта РАН

Для поиска механизмов реализации белок-нуклеинового узнавания предложена автоматическая классификация аминокислот (четкая и с разными типами размытости) на основе анализа геометрических характеристик структур комплексов белок-ДНК. Пространственные контакты между аминокислотами белков и нуклеотидами ДНК определяют при помощи разбиения Вороного-Делоне. Классификацию аминокислот проводят по признакам числа контактов и площадей контактов аминокислот с нуклеотидами. Для четкой классификации аминокислот использованы классические иерархические и неиерархические методы и разные типы расстояний, а для размытой классификации — общий вариационный подход. При помощи предложенной математической модели показано, что около 30% всех контактов между аминокислотами и нуклеотидами в комплексах белок-ДНК являются неслучайными. Методами четкой классификации показано существование инвариантов кластеризации аминокислот. При увеличении числа классов размытой классификации более 6 возникают классы с очень близкими эталонами и равным отнесением объектов к ним, что вводит естественное ограничение на число классов. Полученные данные размытой классификации аминокислот планируется использовать для построения матрицы замен для ДНК-связывающих белков [1].

Работа поддержана грантами РФФИ № 12-07-00634-а и № 14-04-00639-а.

- [1] *Кузнецов Е. Н., Анашкина А. А., Есипова Н. Г., Туманян В. Г.* Кластер-анализ пространственных контактов аминокислотных остатков белков с нуклеотидами ДНК // Машинное обучение и анализ данных, 2014 (в печати).

Cluster analysis of spatial contacts of amino acid residues of proteins with DNA nucleotides

*Kuznetsov Eugene*¹

ken@ipu.ru

*Anashkina Anastasia*²★

nastya@eimb.ru

*Esipova Natalia*²

nge@eimb.ru

*Tumanyan Vladimir*²

tuman@eimb.ru

¹Moscow, Trapeznikov Institute of Control Sciences RAS

²Moscow, Engelhardt Institute of Molecular Biology RAS

For the purpose of protein-DNA recognition analysis, amino acids are classified on the basis of protein-DNA contacts geometry and statistics. However, the methods of crisp classification do not allow describing the diversity of properties of amino acids. Amino acid residues have a variety of properties and can simultaneously belong to different classes. So, the classification of amino acids with different types of fuzzing was used. Voronoi–Delaunay tessellation was used to determine the spatial relationship between the amino acids of proteins and DNA nucleotides. Classification of amino acids was carried out on the statistics of contacts and the statistics of area of contact between amino acids and nucleotides. Classic hierarchical and nonhierarchical methods were used for crisp classification of amino acids with different types of distance measures. General variation approach was used for fuzzy classification of amino acids. Using the proposed mathematical model, it was shown that about 30% of all contacts between amino acids and nucleotides in protein-DNA complexes are not random. Crisp classification methods showed the existence of clustering invariants of amino acids. By fuzzy classification methods, it was shown that 6 classes are optimal for protein-DNA recognition task.

The fuzzy classification of amino acids data is supposed to be used to construct the substitution matrix for DNA-binding protein sequences [1].

This research is funded by RFBR, grants 12-07-00634-a and 14-04-00639-a.

- [1] Kuznetsov E. N., Anashkina A. A., Esipova N. G., Tumanyan V. G. 2014 (in press). Cluster analysis of spatial contacts of amino acid residues of proteins with DNA nucleotides. *Machine Learning and Data Processing*.

Оптимизированная модель для конструирования эффективных малых РНК, играющих важную роль в регуляции работы генов у животных и растений

*Матвеева Ольга Вячеславовна*¹ olga.matveeva@gmail.com
Назипова Нафиса Наиловна^{2*} nnn@impb.psn.ru
*Огурцов Алексей Юрьевич*³ ogurtsov@ncbi.nlm.nih.gov
Шабалина Светлана Алексеевна^{2,3} shabalin@ncbi.nlm.nih.gov

¹Солт-Лейк-Сити, США, Университет штата Юта

²Пушино, Россия, Институт математических проблем биологии

³Бетезда, США, Национальный центр биотехнологической информации

В клеточной биологии искусственные короткие шпилькообразные РНК (кшРНК или shRNAs) стали ценным исследовательским инструментом. Надежный рациональный дизайн экспериментально активных кшРНК важен для дальнейшего функционального изучения генома. Была проанализирована база данных 18 719 молекул кшРНК с известной активностью. В результате поэтапного уменьшения входных параметров модели, являющимися структурными и термодинамическими характеристиками последовательностей кшРНК, из 150 параметров осталось 41. Все эти характеристики-параметры значимо коррелировали с эффективностью кшРНК. В дальнейшем эти параметры были использованы в построении модели методом множественной линейной регрессии. Для тренировки модели было использовано 4/5 базы данных, а для тестирования — 1/5 базы данных. Пятикратная перекрестная проверка модели показала важность параметров, связанных с термодинамическими характеристиками кшРНК ($R = 0,53$). Площадь под ROC (receiver operating characteristic) кривой оказалась равной 0,882, что доказывает пригодность модели для предсказания эффективных кшРНК. На основе оптимизированной модели было разработано программное обеспечение miR_Scan для рационального дизайна эффективных кшРНК [1].

Работа поддержана грантом РФФИ № 12-07-00530.

- [1] *Matveeva O., Nazipova N., Ogurtsov A., Shabalina S.* Optimised models for design of efficient miR30-based shRNAs // *Frontiers in Genetics*, 2012. Vol. 3. Article No. 163. <http://journal.frontiersin.org/Journal/10.3389/fgene.2012.00163/full>.

Optimized model for the design of effective small RNAs that play an important role in regulation of gene expression in animals and plants

*Matveeva Olga*¹

olga.matveeva@gmail.com

Nazipova Nafisa^{2*}

nnn@impb.psn.ru

*Ogurtsov Aleksey*³

ogurtsov@ncbi.nlm.nih.gov

Shabalina Svetlana^{3,2}

shabalin@ncbi.nlm.nih.gov

¹Salt Lake City, USA, University of Utah

²Pushchino, Russia, IMPB RAS

³Bethesda, USA, NCBI, NLM, NIH

Noncoding small hairpin RNAs (shRNAs) became an important research tool in cell biology. Reliable design of these molecules is essential for the needs of large functional genomics projects. An analysis of 18 719 shRNAs with known functional efficiencies has been performed. As a result of gradual reduction of the set of model parameters, which consisted initially of 150 structural and thermodynamic features of shRNA, 41 of them remained which significantly correlate with silencing efficiency. Multiple linear regression analysis was performed using 4/5 of the data for training purposes and the remaining 1/5 for tests. A model which, in addition to the nucleotide preferences, included thermodynamic shRNA features was performing better ($R = 0.53$). As a part of model, 5-fold cross-validation tests of ROC-analysis was performed using binary definition of efficiency for experimental data and continuous model predicted scores for theoretical values. Total area under the curve (0.882) indicates that the developed model is predictive. Software miR_Scan was developed based upon the optimized models [1].

This research is funded by RFBR, grant 12-07-00530.

- [1] Matveeva O., Nazipova N., Ogurtsov A., Shabalina S. 2012. Optimised models for design of efficient miR30-based shRNAs. *Frontiers in Genetics* 3. Article No. 163. <http://journal.frontiersin.org/Journal/10.3389/fgene.2012.00163/full>.

Быстрое распознавание повторов в генетических последовательностях на основе спектрально-аналитического метода

*Панкратов Антон Николаевич**

pan@impb.ru

Пятков Максим Иванович

mpyatkov@gmail.com

Руднев Владимир Рэмович

volodyarv@mail.ru

Куликова Людмила Ивановна

kulikova@impb.ru

Пушино, Институт математических проблем биологии РАН

Разрабатывается метод нахождения различных типов повторов в текстах или кривых, например последовательностях ДНК, пространственных структурах белков и т. д. Метод основан на спектральном анализе функций, которые служат представлением более высокого порядка для текста или кривой, например, GC- и GA-состав для ДНК, кривизна и кручение для пространственной кривой. Результирующие кривые разбиваются на перекрывающиеся фрагменты длины W с шагом d . После этого производится попарное сравнение фрагментов f_i, g_i на основе интегрального оценивания:

$$\rho = \sum_{i=1}^W (f_i - g_i)^2.$$

Для распознавания повторов используется следующее решающее правило: если $\rho < \varepsilon$, где ε — пороговое значение, то фрагменты считаются похожими, а если $\rho \geq \varepsilon$, фрагменты не похожи.

Существенные свойства метода, определяющие его эффективность, следующие:

- интегральное оценивание повторов, которое позволяет не сосредотачиваться на локальных особенностях сигнала;
- масштабирование, т. е. выбор масштаба за счет изменения размеров окна и его шага;
- выполнение операций в пространстве коэффициентов разложения.

Работа поддержана грантами РФФИ № 14-07-00924, 14-07-31306 и 14-07-31196.

- [1] *Pyatkov M. I., Pankratov A. N.* SBARS: Fast creation of dotplots for DNA sequences on different scales using GA-, GC-content // *Bioinformatics*, 2014. Vol. 30. No. 12. P. 1765–1766.

Rapid recognition of repeats in genetic sequences based on the spectral-analytical method

*Pankratov Anton**

pan@impb.ru

Pyatkov Maxim

mpyatkov@gmail.com

Rudnev Vladimir

volodyarv@mail.ru

Kulikova Lyudmila

kulikova@impb.ru

Pushchino, Institute of Mathematical Problems of Biology RAS

A method for finding different types of repeats in text or curves, for example, DNA sequences, protein spatial structures, etc. is under development. The method is based on the spectral analysis of functions which serve as a higher order representation of the sequence or curve, for example, GA- and GC-content curves for DNA, curvature and torsion curves for spatial curves. The resulting curves are divided into overlapping frames with width W and step d . After that, pairwise comparison of frames f_i, g_i has been performed by integral estimation of distance between them:

$$\rho = \sum_{i=1}^W (f_i - g_i)^2.$$

For recognition of repeats, the following decision rule is used: if $\rho < \varepsilon$ where ε is the threshold, then the fragments are considered to be similar; if $\rho \geq \varepsilon$, the fragments are not similar [1].

The essential features of the method that determine its efficiency are the following:

- integral evaluation of repeats, which allows one not to focus on local features of the sequence;
- scaling, i. e., choice of scale by varying the size and step of sliding windows; and
- operations performance in the domain of expansion coefficients.

This research is funded by RFBR, grants 14-07-00924, 14-07-31306, and 14-07-31196.

- [1] Pyatkov M. I., Pankratov A. N. 2014. SBARS: Fast creation of dotplots for DNA sequences on different scales using GA-, GC-content. *Bioinformatics* 30(12):1765–1766.

Применение регуляризованной байесовской логистической регрессии в задаче кредитного скоринга

*Боровых Наталья Игоревна**

natali.bor1992@mail.ru

Красоткина Ольга Вячеславовна

o.v.krasotkina@yandex.ru

Тула, Тульский государственный университет

Несколько последних десятилетий различные методы кредитного скоринга вызывают огромный интерес в банковской отрасли. Кредитный скоринг — это процесс оценивания соискателя на получение кредита, цель которого — определить, является ли потенциальный клиент надежным или нет на основе имеющейся информации о нем и исторических данных о предыдущих клиентах банка. В общем случае оценивается суммарное количество баллов (score) заемщика, набранных по совокупности информативных признаков. Учитываются, например, такие признаки, как возраст, стаж, средний доход семьи и др. Фактически необходимо произвести классификацию заемщиков на два класса на основе подсчитанных баллов.

В данной работе для представления задачи классификации заемщиков выбрана модель байесовской логистической регрессии. В кредитном скоринге, как и во многих других задачах машинного обучения, встает необходимость борьбы с переобучением, поэтому особое внимание в работе уделено исследованию и сравнению эффективности в контексте решаемой задачи различных встроженных методов отбора признакововой информации: гребневой регрессии, лассо Тибширани и модели логистической регрессии с регулируемой селективностью [1].

Работа поддержана грантом РФФИ № 14-07-00964

- [1] *Боровых Н. И., Красоткина О. В.* Применение регуляризованной байесовской логистической регрессии в задаче кредитного скоринга // Известия Тульского гос. ун-та. Технические науки, 2014 (в печати).

Application of the regularized bayessian logistic regression to the credit scoring problem

*Borovykh Natalia**

natali.bor1992@mail.ru

Krasotkina Olga

o.v.krasotkina@yandex.ru

Tula, Tula State University

Several last decades, banks have been rather interested in different credit scoring methods. Credit scoring is a process of assessment of a potential client. The goal of this process is to decide whether the customer is reliable or not based on the data about the client as well as the historical data about the previous customers of the bank. In general case, the sum of the client's score, obtained on the base of the informative features set, is assessed. Such features as age, seniority, average family income, etc. can be taken into account. In fact, the classification in two classes should be performed.

In this paper, the bayessian logistic regression model has been chosen for the problem representation. For struggling with overfitting, the different methods of the feature information selection (ridge regression, Lasso and logistic regression model with adjustable selectivity) are investigated and compared [1].

This research is funded by RFBR, grant 14-07-00964.

- [1] Borovykh N. I., Krasotkina O. V. 2014 (in press). Applying the regularized bayessian logistic regression to the credit scoring problem. *Tula State University Proceedings. Engineering Sciences*.

Алгоритмы непараметрического прогнозирования объемов грузоперевозок

<i>Вальков Антон Сергеевич</i> ¹	valkov@forecsys.ru
<i>Журавлев Юрий Иванович</i> ¹	zhur@ccas.ru
<i>Каширин Даниил Олегович</i> ¹	kashirin@forecsys.ru
<i>Рудаков Константин Владимирович</i> ¹	rudakov@ccas.ru
<i>Стрижов Вадим Викторович</i> ^{1*}	strijov@gmail.com
<i>Корчагин Александр Дмитриевич</i> ²	innovcentre.rzd@mail.ru

¹Москва, ВЦ РАН

²Москва, ЦИР ОАО РЖД

Предмет прогноза — дорожная шахматка. Шахматка — это матрица корреспонденций для прогнозирования элементов матрицы, изменения вагонопотоков с дороги на дорогу. В частности, для устранения тренда и сезонной компоненты предлагается провести сравнение стандартных методов: модель Брауна, модель Хольта–Винтерса, модель авторегрессии — скользящего среднего и ее производные.

Для анализа подходов построения прогностических моделей в работах В. В. Стрижова за 2013 г. приводится обзор непараметрических методов прогнозирования плотности распределения. Используются результаты, полученные Ю. И. Журавлевым в 2002 г. при исследовании задач и построении логических процедур распознавания. Проводится сравнение методов байесовской векторной авторегрессии, векторной авторегрессии с расширенным множеством факторов и динамической стохастической модели общего экономического равновесия, введенной Вальковым в 2012 г. Приводится обзор методов комбинирования моделей прогнозирования плотности с учетом дополнительных ограничений при решении задач классификации, исследованных К. В. Рудаковым (результаты опубликованы в 2005 г) [1].

- [1] *Мотренко А. П., Стрижов В. В.* Построение агрегированных прогнозов объемов железнодорожных грузоперевозок с использованием расстояния Кулльбака–Лейблера // Информатика и её применения, 2104. Т. 8. Вып. 2. С. 86–97.

Nonparametric algorithms for forecasting cargo transportation quantities

*Valkov Anton*¹

valkov@forecsys.ru

*Zhuravlev Yuri*¹

zhur@ccas.ru

*Kashirin Daniil*¹

kashirin@forecsys.ru

*Rudakov Konstantin*¹

rudakov@ccas.ru

Strijov Vadim^{1*}

strijov@gmail.com

*Korchagin Alexandr*²

innovcentre.rzd@mail.ru

¹Moscow, CCAS

²Moscow, CID JSC Russian Railways

This paper investigates a road chessmap forecasting problem. This chessmap is a matrix of correspondence between railway stations that should forecast the matrix elements and measure road to road traffic car changes. To remove a trend and a seasonal component, a comparison of standard methods such as Brown model, Holt–Winters model, autoregressive moving average model and its derivatives is suggested to be made.

For analyzing approaches to construction of forecasting models in the papers by V. Strijov in 2013, here, an overview of nonparametric methods for distribution density estimation is given. The results obtained by Yu. Zhuravlev in 2002 during tasks research and construction of logical recognition procedures have been used. The methods of the Bayesian vector autoregression, vector autoregression with an extended set of factors, and dynamic stochastic model of general economic balance introduced by A. Valkov in 2012 have been compared. Also, the methods for combining density estimation models with regard to additional constraints in solving classification tasks, researched by K. Rudakov (whose results were published in 2005) have been overviewed [1].

- [1] Motrenko A. P., Strijov V. V. 2014. Obtaining an aggregated forecast of railway freight transportation using Kullback-Leibler distance. *Informatika i ee Primeneniya* — *Inform. Appl.* 8(2):86–97.

Прогнозирование объемов грузоперевозок с учетом экзогенных факторов

<i>Вальков Антон Сергеевич</i> ¹	valkov@forecsys.ru
<i>Журавлев Юрий Иванович</i> ¹	zhur@ccas.ru
<i>Каширин Даниил Олегович</i> ^{1*}	kashirin@forecsys.ru
<i>Рудаков Константин Владимирович</i> ¹	rudakov@ccas.ru
<i>Стрижов Вадим Викторович</i> ¹	strijov@gmail.com
<i>Корчагин Александр Дмитриевич</i> ²	innovcentre.rzd@mail.ru

¹Москва, ВЦ РАН

²Москва, ЦИР ОАО РЖД

Исследуются временные методы прогнозирования, числа вагонов определенного типа, прибывающих на станцию в определенный промежуток времени. Предполагается, что измерения числа вагонов производятся через равномерные промежутки, существует календарная сезонность и аддитивный тренд, связанный с влиянием внешних факторов.

Тренд и сезонная составляющая присутствуют только на большом масштабе времени. Вследствие этого на большом масштабе времени предлагается построить параметрическую модель с учетом тренда и сезонности, на малом масштабе предлагается построить непараметрическую модель прогнозирования плотности распределения с экспертно заданной функцией ошибки.

Разработанный в работе А. С. Валькова в 2013 г. метод развивает идею гистограммного прогнозирования с несимметричной функцией потерь. Для построения гистограммы отрезка временного ряда используются методы временного и пространственного агрегирования временных рядов, предложенные В. В. Стрижовым в 2013 г. Значение, минимизирующее свертку гистограммы и функции ошибки, считается оптимальным прогнозируемым значением. Рассматриваются функции ошибки в задачах распознавания, исследованные в работах Ю. И. Журавлёва в 1996 г. и К. В. Рудакова в 1999 г. Предлагается метод выбора наилучшей [1].

- [1] *Вальков А. С., Кожанов Е. М., Мотренко А. П., Хусаинов Ф. И.* Построение кросс-корреляционных зависимостей при прогнозе загрузки железнодорожного узла // *Машинное обучение и анализ данных*, 2013. Т. 1. № 5. С. 505–518.

Forecasting cargo transportation quantities with regard to exogenous factors

Valkov Anton¹

valkov@forecsys.ru

Zhuravlev Yuri¹

zhur@ccas.ru

Kashirin Daniil¹*

kashirin@forecsys.ru

Rudakov Konstantin¹

rudakov@ccas.ru

Strijov Vadim¹

strijov@gmail.com

Korchagin Alexandr²

innovcentre.rzd@mail.ru

¹Moscow, CCAS

²Moscow, CID JSC Russian Railways

The paper analyses time series corresponding to the number of rail-road cars of specific type. The cars arrive at a station in a certain certain time period. Each time series has the following properties: (i) the values of time series are natural numbers; (ii) the time ticks have a fixed interval; (iii) the time series have the calendar seasonality; and (iv) the time series have an additive trend related to the effects of external factors.

Specified problems of the trend and seasonality are held only on a large time scale. Therefore, on a large time scale, the paper proposes a parametric model taking into account the trend and the seasonality. On a small time scale, the paper proposes a nonparametric model for distribution density estimation with expert-specified error function.

The method proposed by A. S. Valkov in 2013 develops an idea of histogram forecasting with an asymmetric loss function. To build a histogram of time series segment, the methods of temporal and spacial time series aggregation have been used. As optimal forecasting value, the value minimizing convolution of histogram and error function has been chosen. Various types of error functions have been considered in solving the recognition task investigated by Yu. I. Zhuravlev in 1996 and K. V. Rudakov in 1999. A method for choosing the best one is suggested [1].

- [1] Valkov A. S., Kozhanov E. M., Motrenko A. P., Husainov F. I. 2013. Constructing a cross-correlation model to forecast the utilization of a railway junction station. *Machine Learning and Data Analysis* 1(5):505–518.

Независимая многовариантная экспертиза в задачах исследования слабо формализованных систем управления

Дорофеев Александр Александрович^{1,2*} daa2@mail.ru
Дорофеев Юлия Александровна^{1,2} dorofeyuk_julia@mail.ru
*Покровская Ирина Вячеславовна*¹ ivp750@mail.ru
*Чернявский Александр Леонидович*¹ achern@ipu.ru

¹Москва, ИПУ РАН

²Москва, НИУ ВШЭ

Разработана модификация метода коллективной многовариантной экспертизы (КМВЭ), адекватная задачам межведомственного типа, названная методом независимой КМВЭ (НМВЭ). Основным отличием разработанного метода НМВЭ от КМВЭ состоит в том, что в процессе экспертизы разрабатываются не варианты решения исходной задачи в целом, а выявляются и разрабатываются варианты решения относительно независимых проблем, связанных с исходной задачей, совокупное решение которых обеспечивает решение исходной задачи. Реализация метода НМВЭ разбивается на 5 основных этапов: формирование списка кандидатов в эксперты; формирование списка относительно независимых проблем экспертизы; оценка компетентности потенциальных экспертов; формирование экспертных комиссий, в том числе для проблем, по которым имеются разногласия; работа сформированных экспертных комиссий. В заключение работы НМВЭ консалтинговая группа формирует итоговый проект решения задачи. Метод НКМВЭ использовался при решении задачи организации и управления рынком межрегиональных автобусных перевозок и задачи создания интеллектуальной системы управления лечебно-диагностическими процессами в крупном медицинском учреждении [1].

Работа поддержана грантами РФФИ № 14-07-00463-а, № 13-07-00992-а, № 13-07-12201-офи и № 12-07-00540-а.

[1] *Дорофеев А. А., Дорофеев Ю. А., Покровская И. В., Чернявский А. Л.* Независимая многовариантная экспертиза в задачах исследования слабо формализованных систем управления // Машинное обучение и анализ данных, 2014 (в печати).

Independent multivariate expertise in the weakly formalized management systems research

Dorofeyuk Alexander^{1,2}★

daa2@mail.ru

Dorofeyuk Yulia^{1,2}

dorofeyuk_julia@mail.ru

*Pokrovskaya Irina*¹

ivp750@mail.ru

*Chernyavskiy Alexander*¹

achern@ipu.ru

¹Moscow, ICS RAS

²Moscow, SIU HSE

The modification of the collective multivariate expertise (CMVE) method was developed, adequate to the interdepartmental type problems, called as the method of an independent CMVE (IMVE). The main and very significant difference between the developed IMVE method from the CMVE one is that in the process of examination, there are developed not solutions of the original problem in general, but solutions to relatively independent problems associated with the initial objective are identified and developed, which comprehensive solution ensures the solution of the original problem. Implementation of the method of IMVE is divided into 5 main stages: creating the list of candidates; the relatively independent list of the expertise problems; the assessment of the potential experts competence; establishment of expert commission including issues on which there are different opinions; and the work of expert commissions. In conclusion, consulting group, which conducted the IMVE, forms the final draft of the problem solution. The IMVE method was used for solving two large-scale applied problems — organization and management of the interregional bus services market problem and the task of creating intelligent control systems for medical-diagnostic processes in the large-scale medical clinic [1].

This research is funded by RFBR, grants 14-07-00463, 13-07-00992, 13-07-12201, and 12-07-00540.

- [1] Dorofeyuk A., Dorofeyuk Y., Pokrovskaya I., Chernyavskiy A. 2014 (in press). Independent multivariate expertise in the weakly formalized management systems research. *Machine Learning and Data Analysis*.

**Комплекс алгоритмов интеллектуального анализа
сложно организованных данных
при исследовании слабо формализованных
систем управления**

Дорофееук Юлия Александровна^{1,2*} dorofeyuk_julia@mail.ru

*Покровская Ирина Вячеславовна*¹ ivp750@mail.ru

*Киселёва Нелли Евсеевна*¹ nellkis@mail.ru

¹Москва, ИПУ РАН

²Москва, НИУ ВШЭ

Рассматривается задача исследования крупномасштабной системы управления, которая состоит из большого числа объектов, каждый из которых характеризуется набором разнородных параметров. Для решения поставленной задачи предлагается исследовать структуру взаиморасположения этих объектов в пространстве информативных параметров. Это позволяет существенно повысить эффективность анализа функционирования системы, а также устойчивость процедур принятия управленческих решений. Для выявления такой структуры разработан специальный комплекс алгоритмов интеллектуального анализа сложно организованных данных, а также процедур экспертной коррекции. Комплекс включает в себя следующие алгоритмы: структурно-классификационного анализа данных (СКАД), выбора информативных параметров, выбора начальных условий для алгоритмов классификации, выбора «оптимального» числа классов, заполнения пропущенных наблюдений, а также процедуры экспертной коррекции результатов работы этих алгоритмов. Проведен теоретический анализ различных вариантов алгоритма СКАД, доказаны теоремы о сходимости алгоритма к локальному экстремуму соответствующего критерия качества.

Работа поддержана грантами РФФИ № 14-07-00463-а, № 13-07-00992-а, № 13-07-12201-офи и № 12-07-00540-а.

[1] *Дорофееук Ю. А., Покровская И. В., Киселёва Н. Е.* Комплекс алгоритмов интеллектуального анализа сложно организованных данных при исследовании слабо формализованных систем управления // Машинное обучение и анализ данных, 2014 (в печати).

The complicated data mining algorithms complex in the study of weakly formalized management systems

Dorofeyuk Yulia^{1,2*}

dorofeyuk_julia@mail.ru

*Pokrovskaya Irina*¹

ivp750@mail.ru

*Kiseleva Nelly*¹

nellkis@mail.ru

¹Moscow, ICS RAS

²Moscow, SIU HSE

The problem of the large-scale management system study is considered. The system consists of a large number of objects, each of which is characterized by a heterogeneous set of parameters. To solve the set tasks, it is proposed to investigate the structure of the relative location of these objects in the informative parameters space. This allows to significantly increase the analysis efficiency of the system functioning and the stability of the procedures for making management decisions. To identify such patterns, special mining complicated data algorithms complex and expert correction procedures were designed. The complex includes the following algorithms: structural-classification data analysis (SCDA), selection of informative parameters, selection of the initial conditions for classification algorithms, selection of the “optimal” number of classes, completing missing observations, as well as procedures of expert correction of the results for these algorithms. The theoretical analysis of different variants of SCDA algorithm was carry out, and the algorithm convergence theorems to the local extremum of the appropriate quality criterion were proved.

This research is funded by RFBR, grants 14-07-00463, 13-07-00992, 13-07-12201, and 12-07-00540.

- [1] Dorofeyuk J. A., Pokrovskaya I. V., Kiseleva N. E. 2014 (in press). The complicated data mining algorithms complex in the study of weakly formalized management systems. *Machine Learning and Data Analysis*.

Вычислимая модель железнодорожных грузоперевозок с учетом коммуникационных ограничений

Шананин Александр Алексеевич^{1,2}

alexshan@yandex.ru

¹Москва, МФТИ ГУ

²Москва, ВЦ РАН

В России наблюдается дефицит финансовых ресурсов, доступных для обновления и модернизации основных фондов естественных монополий. В качестве возможного подхода к решению проблемы обычно предлагается выделить некоторый сегмент деятельности естественной монополии и привлечь в него частных инвесторов, предоставив им преференции по извлечению коммерческой прибыли. В российской системе железнодорожных перевозок обновление вагонного парка сопровождалось привлечением частных инвесторов, допуском их на рынок грузоперевозок. Естественным языком для анализа результатов такого реформирования системы железнодорожных грузоперевозок являются математические модели, учитывающие специфику экономических отношений, сложившихся в отрасли. В докладе рассматривается модель конкурентного равновесия, разработанная для анализа проблем формирования тарифной и инвестиционной политики управления железнодорожными грузоперевозками в современных российских условиях. Получен вариационный принцип в форме пары взаимно двойственных задач выпуклого программирования, из решения которых находится конкурентное равновесие. Идентификация модели по данным о железнодорожных грузоперевозках за 2013 г. выявила наличие несовершенной конкуренции, которая может быть обусловлена коммуникационными ограничениями. Анализ влияния коммуникационных ограничений основывается на исследовании обратной задачи (проблемы моментов) и связанных с ней комбинаторных структур [1].

Работа поддержана грантом РФФИ №13-07-13110 офи-м-РЖД.

- [1] *Ващенко М. П., Гасников А. В., Молчанов Е. Г., Поспелова Л. Я., Шананин А. А.* Вычислимые модели и численные методы для анализа тарифной политики железнодорожных перевозок // *Сообщения по прикладной математике*, 2014.

Rail freight computational model under communication constraints.

Shananin Alexander^{1,2}

alexshan@yandex.ru

¹Moscow, MIPT

²Moscow, CCAS

In Russia, there is a shortage of financial resources available for the renewal and modernization of fixed assets of natural monopolies. Cutting a certain segment of the natural monopoly activities and attracting private investors into it giving them preference for extracting profits are usually proposed as a possible solution for this problem. In the Russian system of rail freightage, car fleet renewal was performed by private investors, which were admitted to the cargo transportation market. Mathematical models taking into account specific economic relations prevailing in the industry are the natural language for analysis of the results of the rail freight system reform. The report examines the competitive equilibrium model developed for the analysis of the problems of tariff and investment policy management of rail freight under modern Russian conditions. The variational principle in the form of a pair of mutually dual convex programming problems, solutions of which is a competitive equilibrium, is obtained. Identification of the model using rail freight data in 2013 revealed the presence of imperfect competition, which may be attributed to communication constraints. Analysis of the communication constraints impact is based on study of the inverse problem (the problem of moments) and related combinatorial structures [1].

This research is funded by RFBR, grant 13-07-13110 ofi-m-RZD.

- [1] Vashchenko M., Gasnikov A., Molchanov E., Pospelova L., Shananin A., 2014 (in press). Computational models and algorithms for the rail freight tariff policy analysis. *Bulletin on Applied Mathematics*.

Классификация фрагментов текста при экспертизе диссертаций на предмет заимствований (плагиата)

*Авдеева Нина Владимировна**

avdeeva@rsl.ru

Лобанова Галина Андреевна

diss@rsl.ru

Никулина Ольга Владимировна

diss@rsl.ru

Москва, Российская государственная библиотека

Российская государственная библиотека оказывает услугу проверки научных текстов на наличие некорректных заимствований (плагиата). Для первичного сопоставления проверяемых работ с текстами из Электронной библиотеки диссертаций и авторефератов (ЭБД РГБ) используется программный комплекс «Антиплагиат.РГБ», способный быстро и эффективно находить совпадающие фрагменты в русскоязычных текстах. Найденные совпадения изучает эксперт, он классифицирует заимствования, выявляет некорректные и оценивает их смысловую значимость. Методологию проверки определяют три фактора: (1) незакрепленность понятия «плагиат» в законодательстве РФ; (2) сосуществование различных подходов к борьбе с плагиатом; (3) специфика диссертационного исследования как объекта авторского права.

Независимость экспертизы, проводимой в РГБ, обусловлена тем, что эксперты не имеют прямого отношения к авторству проверяемых работ, беспристрастны к работам, выполненным в традициях конкретной научной школы, не зависят в своих выводах от денежной стоимости заказа, не участвуют в выборе заказа на проверку и не находятся под влиянием заинтересованных лиц. При этом они формируют свое заключение, опираясь на объективные факты: произведенный системой «Антиплагиат.РГБ» отбор источников, наличие текстовых совпадений и др. [1].

Работа поддержана грантом РФФИ № 13-07-00975.

- [1] *Авдеева Н. В. Лобанова Г. А.* Классификация фрагментов текста при экспертизе диссертаций на предмет заимствований (плагиата) // Информационные ресурсы России, 2014. № 1. С. 2–6. http://www.aselibrary.ru/digital_resources/digital_resources69/digital_resources6970/digital_resources69705071/.

Classification of text fragments in the examination of theses on the subject of borrowing (plagiarism)

*Avdeeva Nina**

avdeeva@rsl.ru

Lobanova Galina

diss@rsl.ru

Nikulina Olga

diss@rsl.ru

Moscow, Russian State Library

The Russian State Library provides a service of testing scientific texts for plagiarisms. For primary matching of the tested paper with those ones of the Digital library of dissertations and author's theses (DDL RSL), there is used a program complex "Antiplagiat.RSL" specially developed for a rapid and effective search of identical fragments in the Russian texts. Afterwards, the found fragments are studied by an expert who classifies the borrowings, detects plagiarisms, and evaluates their scientific significance for the paper. The methodology of this test is determined by three factors: (i) absence of term "plagiarism" in RF legislation; (ii) existence of diverse practices of fighting plagiarism; and (iii) specificity of dissertation research as of a Copyright unit.

The plagiarism test carried out at the RSL is independent in its essence, as the RSL experts are not the interested parties: they are not connected with authorship of the paper; they have no preferences to ideas of any scientific school; their opinion does not depend on the price of the order, and they do not participate in choosing the paper for testing. Finally, they are not under influence of certain interested parties. Meanwhile, they ground their opinions only on the facts, i. e., on the sources selected by the "Antiplagiat.RSL" system, on the revealed coincident places in the texts, etc. [1].

This research is funded by RFBR, grant 13-07-00975.

- [1] *Avdeeva N., Lobanova G.* 2014. Classification of text fragments in the examination of theses on the subject of borrowing (plagiarism). *Information Resources of Russia* 1:2-6. http://www.aselibrary.ru/digital_resources/digital_resources69/digital_resources6970/digital_resources69705071/.

Расчет семантической близости концептов на основе кратчайших путей в графе ссылок Википедии

*Варламов Максим Игоревич**

varlamov@ispras.ru

Коршунов Антон Викторович

korshunov@ispras.ru

Москва, Институт системного программирования РАН

Семантическая близость пары понятий, или концептов, — мера, отражающая степень их смысловой связанности. В системе анализа текстов Текстерра (ИСП РАН) мера семантической близости (мера Дайса) используется для разрешения лексической многозначности терминов. Поскольку база знаний Текстерры строится на основе Википедии с сохранением графа гиперссылок между статьями (концепты в Текстерре — статьи Википедии), мы заинтересованы в точных ссылочных мерах семантической близости. Локальные ссылочные меры, рассматривающие только ближайшие окрестности вершин, не могут дать оценку близости для произвольной пары концептов. Недостатком же глобальных мер, использующих весь граф, является слишком высокая временная сложность.

В представленной работе предлагается глобальная мера семантической близости концептов на основе расстояния (длины кратчайшего пути) между ними в графе Википедии.

Отличие предлагаемого метода от существующих мер на основе кратчайших путей заключается в использовании знаний о неравноценности различных типов ссылок Википедии для решения данной задачи. Кроме того, для расчетов используются специализированные структуры данных, позволяющие эффективно вычислять расстояния между вершинами с приемлемыми расходами по памяти.

В сравнении с мерой Дайса показывается, что использование кратчайших путей позволяет как повысить корреляцию получаемых оценок близости с экспертными, так и достичь лучших результатов в задаче разрешения лексической многозначности [1].

- [1] *Варламов М. И., Коршунов А. В.* Расчет семантической близости концептов на основе кратчайших путей в графе ссылок Википедии // Машинное обучение и анализ данных, 2014 (в печати).

Computing semantic similarity of concepts using shortest paths in Wikipedia link graph

Varlamov Maksim

varlamov@ispras.ru

Korshunov Anton

korshunov@ispras.ru

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

A measure of semantic similarity between concepts characterizes the degree of relatedness between their senses. Texterra, a system for text analysis from ISP RAS, uses Wikipedia-based Dice semantic similarity measure for word sense disambiguation. Since Texterra's knowledge base is built from Wikipedia preserving structure of its hyperlink graph (Texterra's concepts are Wikipedia's articles), we are interested in precise link-based semantic similarity measures. Local measures use only nearest neighborhood of vertices and thus cannot provide the value of similarity for an arbitrary pair of concepts. Global measures use the entire graph and thus have too high time complexity.

This work presents a global semantic similarity measure based on distances (lengths of shortest paths) between concepts in Wikipedia link graph.

The difference of the proposed method from existing measures based on the shortest paths is in the usage of disparity of different link types. Additionally, the present authors use a special data structure which allows to compute the shortest paths efficiently with acceptable memory costs.

Comparison the present method with Dice measure used in Texterra system shows that the use of the shortest paths allows both to increase the correlation between computed and expert similarity and to achieve better results in the word sense disambiguation task [1].

- [1] Varlamov M., Korshunov A. 2014 (in press). Computing semantic similarity of concepts using shortest paths in Wikipedia link graph. *Machine Learning and Data Analysis*.

Многокритериальные и многомодальные вероятностные тематические модели коллекций текстовых документов

Воронцов Константин Вячеславович^{1,2,3,4}★ voron@forecsys.ru

Потапенко Анна Александровна^{2,4} anya_potapenko@mail.ru

*Фрей Александр Ильич*¹ oleksandr.frei@gmail.com

*Апишев Мурат Азаматович*³ great-mel@yandex.ru

*Дойков Никита Владимирович*³ nikitad101@gmail.com

*Шапулин Андрей Валентинович*³ a.shapulin@gmail.com

*Чиркова Надежда Александровна*³ nadiinchi@gmail.com

Москва, ¹МФТИ, ²ВЦ РАН, ³ВМК МГУ, ⁴НИУ ВШЭ

Вероятностная тематическая модель выявляет тематику коллекции текстовых документов, описывая темы дискретными распределениями на множестве терминов, документы — дискретными распределениями на множестве тем. *Аддитивная регуляризация тематических моделей (ARTM)* основана на максимизации взвешенной суммы логарифма правдоподобия и дополнительных критериев — регуляризаторов. Исследуются комбинации регуляризаторов сглаживания, разреживания, декоррелирования и отбора тем, совместно повышающих интерпретируемость тем. Предлагается способ построения многомодальных тематических моделей, в которых документы связываются не только со словами, но и с категориями, метками времени, авторами, пользователями. В докладе рассматриваются применения ARTM для моделирования различных текстовых коллекций: иерархическая модель тематики конференций ММРО и ИОИ за ряд лет; динамическая модель потока пресс-релизов высших органов государственной власти и органов внешних сношений ряда стран; модель классификации символьных последовательностей для диагностики многих заболеваний по дискретизированному электрокардио-сигналу. Затрагиваются вопросы реализации параллельной распределённой библиотеки BigARTM для многокритериального тематического моделирования больших текстовых коллекций.

Гранты РФФИ: 14-07-00847, 14-07-00908, 14-07-31176.

- [1] Vorontsov K. V., Potapenko A. A. Additive Regularization of Topic Models // Machine Learning Journal, Special Issue on Data Analysis and Intelligent Optimization, Springer, 2014 (to appear).

Multi-criteria and multi-modal probabilistic topic models of text collections

Vorontsov Konstantin^{1,2,3,4}

voron@forecsys.ru

Potapenko Anna^{2,4}

anya_potapenko@mail.ru

*Frei Olexandr*¹

oleksandr.frei@gmail.com

*Apishev Murat*³

great-mel@yandex.ru

*Doikov Nikita*³

nikitad101@gmail.com

*Shapulin Andrey*³

a.shapulin@gmail.com

*Chirkova Nadezhda*³

nadiinchi@gmail.com

Moscow, ¹MIPT, ²CC RAS, ³MSU, ⁴HSE

Probabilistic Topic Modeling (PTM) is a rapidly developing statistical tool for revealing a hidden thematic structure of text collections. PTM describes each topic by a multinomial distribution over words and represents each document by a multinomial distribution over topics. Learning PTM is an ill-posed problem of stochastic matrix factorization. We develop a non-Bayesian approach of *Additive Regularization of Topic Models* (ARTM) based on the maximization of a weighted sum of the log-likelihood and additional criteria. We show that a combination of sparsening, smoothing, decorrelation and topic selection regularizers improves the interpretability of topics with almost no loss of the likelihood. Then we propose a multi-modal PTM, in which documents can be linked with authors, class labels, categories, time stamps, users, etc.

We report several applications of ARTM: hierarchical model of Russian conferences Mathematical Methods of Pattern Recognition (MMPO) and Intelligent Information Processing (ИОИ) for 7 years; dynamic model of official government documents from several countries for 10 years; classification of symbolic sequences for the diagnosis of many diseases from a discretized ECG-signal. Also we consider the implementation of distributed parallel multi-criteria and multi-modal PTMs in the open source library BigARTM for processing large text collections.

Thanks to RFBR grants: 14-07-00847, 14-07-00908, 14-07-31176.

- [1] *Vorontsov K. V., Potapenko A. A.* Additive Regularization of Topic Models // Machine Learning Journal, Special Issue on Data Analysis and Intelligent Optimization, Springer, 2014 (to appear).

Структурные аналогии в символьных последовательностях различной языковой природы

Гусев Владимир Дмитриевич

`gusev@math.nsc.ru`

*Мирошниченко Любовь Александровна**

`luba@math.nsc.ru`

Саломатина Наталья Васильевна

`nataly@math.nsc.ru`

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

Символьные последовательности (тексты) являются объектом исследования в математике, информатике, биологии, лингвистике, музыке. Изучение структуры текстов позволяет решать важные классификационные, эволюционные и дешифровочные задачи в указанных областях знания. При всем многообразии текстов их объединяет понятие повтора в широком смысле как элементарного структурообразующего элемента. Целью работы является типизация повторов и их комбинаций, т. е. структурных единиц более высокого уровня. Обсуждается их функциональная нагрузка, проводятся межъязыковые аналогии. Представление о наиболее характерных структурных единицах дают фрагменты текста с аномально низкой (по Колмогорову) сложностью. Для их обнаружения используются сложностные профили последовательности (введены авторами) и аппарат сканирующих статистик (адаптирован авторами для текстов на естественном языке). Отличительной особенностью используемого подхода является применимость его к различным языковым системам. При отборе структур для иллюстрации подхода авторы руководствовались соображениями «межъязыковой общности». Межъязыковые аналогии на уровне структур могут инициировать перенос идей и понятий, связанных с одной языковой системой, на другую, включая постановку содержательных задач и способы их решения [1].

Работа поддержана грантом РФФИ № 13-07-00400.

- [1] *Гусев В. Д., Мирошниченко Л. А., Саломатина Н. В.* Структурные аналогии в символьных последовательностях различной языковой природы // Машинное обучение и анализ данных, 2014 (в печати).

Structural analogies in symbolic sequences of different nature

Gusev Vladimir

`gusev@math.nsc.ru`

*Miroshnichenko Liubov**

`luba@math.nsc.ru`

Salomatina Natalia

`nataly@math.nsc.ru`

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

Symbolic sequences (words, strings, texts) as an object of study are encountered in various areas of knowledge: mathematics, informatics, biology, linguistics, and music. Many problems related to analysis of texts are of classification, evolutionary, and deciphering nature. Despite of diversity of texts, the existence of repetitions as elementary structure-forming units is general for them. The purpose of the work is typification of repetitions and their combinations, i. e., structural units of higher level. The complexity profiles of sequence (introduced by the authors) and scanning statistics (adapted for texts in a natural language) are used for revealing such structures. As a result of processing of different texts, the structural units which are characterized by their “interlingual” nature are allocated and described.

This research is funded by RFBR, grant 13-07-00400.

- [1] Gusev V. D., Miroshnichenko L. A., Salomatina N. V. 2014 (in press). Structural analogies in symbolic sequences of different nature. *Machine Learning and Data Analysis*.

Экспериментальное исследование методов выявления нечетких дубликатов научных публикаций

Дербенёв Николай Викторович nicvic@mail.ru
Козлюк Дмитрий Александрович dmitry.kozliuk@gmail.com
*Никитин Вадим Владимирович** vadimirtoej@gmail.com
Толчеев Владимир Олегович tolcheevvo@mail.ru

Москва, Национальный исследовательский университет «МЭИ»

Рассмотрены известные методы, в том числе авторский — обобщенный коэффициент ассоциативности (ОКА), для выявления нечетких дубликатов в научных публикациях и критерии эффективности их работы. Сформулирован целевой критерий работы методов, сочетающий требования к полноте и к точности. Составлена выборка пар документов, представленных библиографическими описаниями (заголовками и аннотациями), получены экспертные оценки схожести документов в парах. Проведены эксперименты по установлению наибольшей точности результатов различных методов при ограничении на полноту. Для коэффициента ассоциативности Джаккарда и ОКА, продемонстрировавших наилучшие результаты, предложены и апробированы способы повышения точности до 74 % при полноте 90 %. Результаты проверены путем анализа полнотекстовых описаний части документов исследуемой выборки, доступных публично.

- [1] *Дербенёв Н. В., Козлюк Д. А., Никитин В. В., Толчеев В. О.* Экспериментальное исследование методов выявления нечетких дубликатов научных публикаций // Машинное обучение и анализ данных, 2014. Т. 1. № 7. С. 875–884. <http://jmla.org/papers/doc/2014/no7/Derbenev2014Experimental.pdf>.

Experimental research of near-duplicate detection methods for scientific papers

Derbenev Nikolay

nicvic@mail.ru

Kozliuk Dmitry

dmitry.kozliuk@gmail.com

*Nikitin Vadim**

vadimirtoej@gmail.com

Tolcheev Vladimir

tolcheevvo@mail.ru

Moscow, Moscow Power Engineering Institute

Near-duplicate detection problem focuses on determining pairs of semantically equivalent documents which differ syntactically. For scientific papers, this case corresponds either to plagiarism among different authors or to a single author publishing “cloned” papers to achieve higher citation rank. Given a set of document pairs with *a priori* expert opinions, efficiency of any near-duplicate detection method can be measured by the number of correct detections (recall) and false-positive count (precision). These metrics cannot be maximized simultaneously – complex criterion is used.

Improvements can be made by choosing a method with the highest precision for a given recall and then preprocessing documents in favor of specifics of the selected algorithm. The authors propose an efficiency criteria limiting both recall and precision. Then, form a sample set of title and abstract pairs is formed, and expert assessments are acquire for them. After that, a sample set is used to evaluate performance of various near-duplicate detection methods subjected to the criteria proposed.

Precision of 74% at 90% recall for Jaccard and generalized (GSC) similarity coefficients appeared to be reachable by removing frequent words of authors’ vocabulary from the documents’ abstracts. Along with Winnowing, GSC scored best in method comparison without preprocessing. The results were checked on a subset of documents with full texts available (about 20% of sample set). Verification confirmed high sustained precision by revealing documents with near-duplicate titles and abstracts to have identical content [1].

- [1] Dernenev N. V., Kozliuk D. A., Nikitin V. V., Tolcheev V. O. 2014. Experimental research of near-duplicate detection methods for scientific papers. *Machine Learning and Data Analysis*. 1(7):875–884. <http://jmla.org/papers/doc/2014/no7/Derbenev2014Experimental.pdf>.

Регуляризация многоязычных тематических моделей

Дударенко Марина Алексеевна

m.dudarenko@gmail.com

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

Москва, Вычислительный центр им. А. А. Дородницына Российской
академии наук

В многоязычном тематическом моделировании для получения согласованных многоязычных тем из текстовых коллекций нужны дополнительные ресурсы, связывающие языки. Большинство многоязычных моделей основываются на модели LDA (latent Dirichlet allocation) и учитывают либо связанные документы, либо двуязычные словари.

В работе [1] предлагается многоязычная модель на основе аддитивной регуляризации тематических моделей, позволяющая одновременно учитывать связи между документами параллельной или сравнимой коллекции и двуязычный словарь. При использовании информации из двуязычного словаря можно рассматривать матрицу переводов как неизменный объект или же настраивать вероятности переводов в каждой теме.

Качество многоязычных моделей оценивается по средней позиции документа-перевода в поисковой выдаче кросс-язычного поиска. В этом случае запросом является документ на одном языке, а поиск производится среди документов другого языка.

Показано, что комбинированный учет слов-переводов из двуязычного словаря и связанных документов улучшает качество кросс-язычного поиска по сравнению с моделями, использующими только один тип информации. При использовании двуязычных словарей оценивание вероятностей переводов, зависящих от темы, еще больше улучшает качество. Для каждого перевода в словаре тематическая модель позволяет указать контекст (набор тем), в котором данный перевод наиболее уместен.

Работа поддержана грантами РФФИ № 14-07-31176 и № 14-07-00908.

[1] *Дударенко М. А.* Регуляризация многоязычных тематических моделей // Машинное обучение и анализ данных, 2014 (в печати).

Multilanguage topic models regularization

Dudarenko Marina^{1,2}

m.dudarenko@gmail.com

¹Moscow, Lomonosov Moscow State University;

²Moscow, Dorodnicyn Computing Centre of RAS

The goal of multilingual topic modeling is to discover multilingual topics from text corpora. To make them aligned across languages, additional information connecting languages is required. Most multilingual models have been developed mainly on the base of LDA (latent Dirichlet allocation) model with usage of either parallel and comparable corpora or bilingual translation dictionaries.

In the work [1], a multilingual model based on the additive regularization of topic models (ARTM) allowing to combine both parallel or comparable corpus and bilingual translation dictionary is proposed. Two approaches to include information from bilingual dictionary are suggested: the first assumes translation matrix to be constant, and the second learns translation probabilities for each topic.

To measure quality of the proposed multilingual models, cross-language search is performed. For each query document in one language, the position of its translation to another language is computed and then average positions across all queries is calculated.

Combining word translations from bilingual dictionary with connected documents improves cross-lingual search compared against models using just one information source. Learning word translation probabilities for bilingual dictionaries improves results even further. This also allows to determine a context (a set of topics) for each translation where this translation is appropriate.

This research is funded by RFBR, grants 14-07-31176 and 14-07-00908.

[1] Dudarenko M. 2014 (in press). Regularization of multilanguage probabilistic topic models. *Machine Learning and Data Mining*.

Формирование единиц представления знаний в задаче их оценки на основе открытых тестов

Емельянов Геннадий Мартинович Gennady.Emelyanov@novsu.ru

*Михайлов Дмитрий Владимирович** mdv74@list.ru

Козлов Александр Павлович caleo@yandex.ru

Великий Новгород, НовГУ

Тестовое задание открытой формы предполагает ответ испытуемого в виде одного или нескольких предложений естественного языка (ЕЯ). Предложена методика автоматизированного формирования единиц знаний, оцениваемых с применением открытых тестов. В ее основе лежит развиваемая авторами концепция ситуации употребления ЕЯ как единицы описания его семантики. Языковой контекст, фиксируемый данной единицей, отражает значимые в ситуации объекты, отношения между ними и их выражения в текстах, эквивалентных по смыслу. Сказанное позволяет учитывать целостный образ ситуации употребления предметно-ограниченного ЕЯ-подмножества для описания отдельного факта предметной области теста. На основе данной концепции решается задача поиска наиболее рационального плана передачи смысла между экспертами и обучаемыми в системе контроля знаний с применением открытых тестов. При этом сам смысл должен быть отражен в максимально компактном объеме текстовых данных, которые и участвуют в оценке близости ответа испытуемого правильному ответу. В работе предлагается решение указанной задачи выделением и классификацией структурных единиц, соответствующих синтагмам и определяющих лексико-синтаксические связи в рамках смыслового эталона ситуации ЕЯ-употребления. При этом найденные смысловые эталоны позволяют минимум в 4 раза сократить объем текста, необходимого для передачи единицы знаний посредством ЕЯ между его носителями без потери полезной смысловой составляющей [1].

Работа поддержана грантом РФФИ № 13-01-00055.

- [1] *Емельянов Г. М., Михайлов Д. В., Козлов А. П.* Формирование единиц представления предметных знаний в задаче их оценки на основе открытых тестов // Машинное обучение и анализ данных, 2014 (в печати).

Forming of the knowledge units' representations to estimate them by open-form tests

Emelyanov Gennady

Gennady.Emelyanov@novsu.ru

*Mikhaylov Dmitry**

mdv74@list.ru

Kozlov Aleksandr

caleo@yandex.ru

Veliky Novgorod, Yaroslav-the-Wise Novgorod State University

An open-form test task in the system for computer-aided testing of knowledge presupposes a trainee's answer in the form of one or several sentences in natural language (NL). The problem considered is the computer-aided formation of knowledge units for their estimation by means of open-form tests. The decision method is based on the developed-by-authors conception of NL-usage situation as the unit of formalized description of text semantics. Language context fixed by this unit reflects significant objects, relations between them, and expression of these relations in the texts equivalent-by-sense. This allows to take proper account of an integral pattern of usage situation for NL subject-oriented subset at description of specific fact of test's subject area. Here, the problem to seek the most rational plan for sense's transfer among experts and trainees which participate in open-form-testing is solved with application of the suggested conception. The sense as a result has to be reflected in a maximum compact volume of text data. These data are what has to be processed at estimation of affinity to the given correct variant for the trainee's answer. This paper represents how to allocate these data by revelation and classification of structural units which define lexical syntactic links and correspond to syntagms concerning a sense standard for NL usage situation. Here, sense standards are found that make it possible to reduce a volume of text necessary for lossless-in-sense transmission of knowledge unit between native speakers by means of NL [1].

This research is funded by RFBR, grant 13-01-00055.

- [1] Emelyanov G., Mikhaylov D., Kozlov A. 2014 (in press). Forming of the knowledge units' representations to estimate them by application of tests of the open form. *Machine Learning and Data Analysis*.

Компьютерное психологическое тестирование на базе портала MultiTest

Жаркова Оксана Станиславовна

umx@bk.ru

Берестнева Ольга Григорьевна

ogb6@yandex.ru

Моисеенко Александр Владимирович

alex@am.tpu.ru

*Марухина Ольга Владимировна**

marukhina@tpu.ru

Томск, Томский политехнический университет

Статья посвящена проблеме создания универсального инструментария для организации компьютерного психологического тестирования на базе портала MultiTest. Портал представляет собой WEB-сервер с многоуровневым пользовательским доступом, что обеспечивает конфиденциальность и безопасность полученной информации. Дано описание структуры портала, возможных групп пользователей (анонимы, тестируемые, тестирующие, редакторы, аналитики, администраторы, суперадминистраторы) и их возможностей. Особое внимание уделено разработанному авторами «Редактору тестов», с помощью которого создание и добавление новых психологических тестов доступно даже пользователям, не имеющим опыта программирования. Инструменты пользователя с правами редактора представляют собой средства редактирования модулей, расположенных в базе данных. В статье представлены два приложения, разработанные на базе портала: (1) профориентационного тестирования для абитуриентов Томского политехнического университета; (2) оценки компетентности IT-специалистов.

Работа поддержана грантом РФФИ № 14-06-00026.

- [1] *Zharkova O. S., Berestneva O. G., Moiseenko A. V., Marukhina O. V.* Psychological computer testing based on multitest portal // World Appl. Sci. J., 2013. Vol. 24. С. 220–224. [http://www.idosi.org/wasj/wasj\(ITMIES\)13/36.pdf](http://www.idosi.org/wasj/wasj(ITMIES)13/36.pdf).

Psychological computer testing based on MultiTest portal

Zharkova Oksana

umx@bk.ru

Berestneva Olga

ogb6@yandex.ru

Moiseenko Aleksandr

alex@am.tpu.ru

*Marukhina Olga**

marukhina@tpu.ru

Tomsk, Tomsk Polytechnic University

The article is devoted to the problem of creating a universal tool for computer psychological testing based on portal MultiTest. The portal is a WEB-server with multiuser access to ensure privacy and security of information received. The description of the portal structure, potential user groups (anonymous, tested, teachers, editors, analysts, administrators, and super administrators) and their capabilities is given. Particular attention is given to the "Test editor" developed by authors, with the help of which creating and adding new psychological tests are available even for users without programming experience. The tools available for user with editor's rights are editing modules arranged in a database. The paper presents two applications developed on the portal basis: (1) a vocational guidance test for applicants of the Tomsk Polytechnic University; and (2) assessing the IT-specialists competence.

This research is funded by RFBR, grant 14-07-14-06-00026.

- [1] Zharkova O., Berestneva O., Moiseenko A., Marukhina O. 2013. Psychological computer testing based on multitest portal. *World Appl. Sci. J.* 24:220–224. [http://www.idosi.org/wasj/wasj\(ITMIES\)13/36.pdf](http://www.idosi.org/wasj/wasj(ITMIES)13/36.pdf).

Способы повышения точности классификации двуязычных текстовых библиографических документов

*Мохов Андрей Сергеевич**

asmokhov@mail.ru

Толчеев Владимир Олегович

tolcheevvo@mail.ru

Москва, НИУ «МЭИ»

Рассматривается задача повышения точности классификации текстов при обработке и анализе двуязычной информации. Проводятся исследования известных процедур классификации и предлагаются два подхода для решения поставленной задачи — разработка модификаций профильных методов и синтез коллективов решающих правил (КРП). Даются рекомендации по составлению КРП в рамках рассматриваемой задачи.

Экспериментальные исследования авторов показали, что использование смешанных выборок, которые содержат терминологическую информацию на русском и английском языках, в большинстве случаев обеспечивает более высокую точность классификации по сравнению с одноязычными выборками.

На основе экспериментальных результатов можно сделать вывод о хороших точностных характеристиках профильных методов — методов, основанных на построении формального объекта (профиля) класса, способного описать остальные элементы класса. Эти методы, в том числе новые, предложенные в данной статье, за счет более эффективного выявления информативных терминов позволяют улучшить точность классификации на смешанных (двуязычных) выборках по сравнению с известными «классическими» процедурами — методом *k*-ближайших соседей и методом центроидов.

Приблизительно одинаковая точность всех профильных методов при их высокой разнородности позволяет объединять эти процедуры в КРП — коллективный классификатор, обладающий более высокой точностью классификации двуязычных документов по сравнению с одиночными классификаторами [1].

- [1] *Мохов А. С., Толчеев В. О.* Разработка методов высокоточной классификации двуязычных текстовых библиографических документов // Информационные технологии, 2014. №5. С. 8–13. http://novtex.ru/IT/it2014/number_05_annot.html#3.

Methods of bilingual text documents classification precision increasing

*Mokhov Andrey**

asmokhov@mail.ru

Tolcheev Vladimir

tolcheevvo@mail.ru

Moscow, NRU “MPEI”

The problem of improving the precision of a bilingual text documents classification is considered. A research of well-known classification methods is provided and two ways for solution of this problem — the development of modifications for profile methods and ensembles of classifiers (EoC) application — are suggested. Recommendations for EoC’s construction in a context of this problem are given

The experiments show that in most cases, using of mixed, bilingual corpus in Russian and English provides the higher precision of classification than monolingual samples.

According to their experimental results, the authors get better classification results with profile classification methods - methods based on a construction of a formal object (profile) which can characterize all other elements of a class. These profile methods, including some new ones which are presented in this work, show better results than “classic” methods (k-Nearest Neighbours, Centroid method) on bilingual corpus. This occurs due to better extraction and weighting of informative terms.

Roughly comparable precision with high diversity of profile classification methods allows one to unite them into ensembles of classifiers, which show even better precision of classification on bilingual text documents than mon classifiers.

Both suggested ways of classification precision improving show better results and efficiency on bilingual text documents compared to classic methods [1].

- [1] Mokhov A., Tolcheev V. 2014. The development of high-precision classification methods for bilingual text documents. *Information Technologies* 5:8–13. http://novtex.ru/IT/it2014/number_05_annot.html#3.

Моделирование вариативности произношения для уменьшения уровня ошибок при распознавании речи

*Чучупал Владимир Яковлевич*¹

chuchuccas.ru

Коренчиков Антон Александрович^{2*}

vermillionmail.ru

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

²Московский государственный университет им М. В. Ломоносова

Анализ ошибок автоматического распознавания, а также модельные эксперименты показывают, что за счет использования корректных вариантов произнесения слов можно примерно вдвое уменьшить число ошибок распознавания. В то же время известные из литературы экспериментальные результаты свидетельствуют о гораздо более скромном выигрыше, достигаемом при использовании моделей вариативности произношения.

В работе оценивается перспективность использования, с точки зрения уменьшения пословной ошибки распознавания, моделей вариативности произношения для задачи распознавания слитной русской речи. Приведена вероятностная модель вариативности произношения, способы оценки ее параметров, реализации вычислений в рамках стандартных процедур поиска при распознавании речи.

Показано, что для речевых корпусов данных, которые содержат читаемый речевой материал, в нормальной обстановке, от людей с нормативным произношением использование явных моделей вариативности произношения малоэффективно. В то же время, для распознавания спонтанной разговорной речи и при несоответствии характеристик тестового речевого материала обучающему методу моделирования вариативности произношения могут быть весьма эффективным (снижая относительный уровень так называемой пословной ошибки распознавания на 17%) способом уменьшения числа ошибок распознавания [1].

Работа поддержана грантом РФФИ № 14-01-00607.

- [1] *Чучупал В. Я., Коренчиков А., А.* Моделирование вариативности произношения для уменьшения уровня ошибок при распознавании речи // Машинное обучение и анализ данных, 2014 (в печати).

Improving speech recognition accuracy by means of word pronunciation modeling

*Chuchupal Vladimir*¹

chuchu@ccas.ru

Korenchikov Anton^{2*}

vermilion@mail.ru

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

²Moscow, Lomonosov Moscow State University

The analysis of automatic speech recognizer errors as well as the data of “oracle” style experiments show that the utilization of the correct pronunciation phonemic transcriptions during the recognition could lower the word error rate (WER) level as much as 50%.

At the same time, the known from the literature experiments on the speech corpora show only a insignificant improvements in WER that are obtained with the implementation of pronunciation variation models.

In this study, the potential, from the point of lowering the WER level, of the pronunciation variation models has been estimated for the task of automatic recognition of fluent natural Russian speech.

The probabilistic pronunciation model is defined as well as the methods for estimation of its parameters and implementation of the multiple word phone transcriptions during the conventional search procedures.

Numerical experiments show that for the conventional speech corpora with read speech material, recorded in the calm environment, with normal speaking rate, and standard pronunciation, the use of the pronunciation modeling techniques could be not effective. At the same time, for the corpora that contain a natural fluent speech, recorded under the conditons that differ from those of the training material, the use of the pronunciation variation models could bring the substantial improvements in WER. In the authors’ experiments, the relative improvements in WER were more than 17% that makes the pronunciation variation models a promising approach [1].

This research is funded by RFBR, grant 14-01-00607.

- [1] Chuchupal V., Korenchikov A. 2014 (in press). Improving speech recognition accuracy by means of word pronunciation modeling. *Machine Learning and Data Analysis*.

Графическая вероятностная модель со скрытыми состояниями на основе главных многообразий

Юлин Сергей Сергеевич^{1*}

julin.serg@gmail.com

*Паламарь Ирина Николаевна*²

irina.palamar@mail.ru

¹Рыбинск, ОАО «КБ"ЛУЧ»

²Рыбинск, РГАТУ имени П. А. Соловьева

Предлагается графическая вероятностная модель на основе аппроксимации обучающих данных главными многообразиями малой размерности, заданными в виде сетки узлов. В качестве алгоритма построения сетки узлов используется алгоритм самоорганизующихся карт Кохонена. Нормированное от 0 до 1 расстояние от наблюдаемых данных до каждого узла сетки может являться аналогом вероятности появления наблюдаемых данных в скрытых состояниях и использоваться в алгоритмах вероятностного вывода на предлагаемой модели. За основу предлагаемой модели взята структура скрытой Марковской модели (НММ — Hidden Markov Model), которая представляет собой граф из двух типов вершин, соответствующих двум типам случайных величин (наблюдаемые и скрытые) и ребер между ними. Каждому скрытому состоянию соответствует узел аппроксимирующей сетки. Обучение модели производится итерационной процедурой, направленной на максимизацию правдоподобия. Вероятностный вывод с целью вычисления правдоподобия производится алгоритмом «прямого-обратного хода». Предлагаемая в работе модель показала лучшие среди тестируемых моделей (НММ, HCRF — Hidden Conditional Random Fields) результаты классификации при обучении на малом количестве обучающих данных (менее 100 экземпляров каждого класса) на тестовых наборах “Character Trajectories Data Set” и “Spoken Arabic Digit Data Set” из репозитория UCI. К недостаткам предлагаемой модели следует отнести значительно большее количество скрытых состояний, чем в моделях НММ и HCRF, что сказывается на увеличении времени выполнения классификации [1].

- [1] Паламарь И. Н., Юлин С. С. Порождающая графическая вероятностная модель на основе главных многообразий // Труды СПИИРАН, 2014. С. 227–247. <http://proceedings.spiiras.nw.ru/ojs/index.php/sp/article/view/1837/1683>.

Probabilistic graphical model with hidden states based on the principal manifolds

Yulin Sergey^{1*}

julin.serg@gmail.com

*Palamar Irina*²

irina.palamar@mail.ru

¹Rybinsk, JSC Lutch

²Rybinsk, P. A. Solovyov Rybinsk State Aviation Technical University

A probabilistic graphical model based on approximation of training data with principal manifolds of small dimension given in the form of grid nodes is suggested. Kohonen's self-organizing map algorithm is used as an algorithm for constructing the grid nodes. Normalized from 0 to 1, the distance from the observed data to each node is considered as an analog of the probability of occurrence of the observed data in the hidden states and is used in probabilistic inference algorithms on the proposed model. The basis of the proposed model is Hidden Markov Model (HMM) structure, which is a graph of two types of nodes corresponding to two types of random values (observed and hidden) and edges between them. Each hidden node corresponds to a node of the approximating grid. Model's learning is performed by an iterative procedure directed to maximizing likelihood. Probabilistic inference to compute the likelihood is performed using "forward-backward" algorithm. The proposed model showed better results than the HMM and HCRF (Hidden Conditional Random Fields) at training on a small amount of training data (less than 100 copies of each class) on the test sets from the UCI repository: "Character Trajectories Data Set" and "Spoken Arabic Digit Data Set." The disadvantage of the proposed model is considerably larger amount of hidden states needed comparing with HMM and HCRF classifiers which leads to longer classification runtimes [1].

- [1] Palamar I., Yulin S. 2014. Generative probabilistic graphical model base on the principal manifolds. *SPIIRAS Proceedings*. 227–247. <http://proceedings.spiiras.nw.ru/ojs/index.php/sp/article/view/1837/1683>.

Методы машинного обучения при восстановлении обстановки осадконакопления в задачах геостатистического моделирования

*Белозеров Борис Владимирович*¹ Belozеров.BV@gazpromneft-ntc.ru

*Бочков Андрей Сергеевич*¹ Bochkov.AS@gazpromneft-ntc.ru

*Урмаев Олег Станиславович*¹ Ushmaev.OS@gazprom-neft.ru

Фукс Ольга Михайловна^{1,2*} Phuks.OM@gazprom-neft.ru

¹ Санкт-Петербург, ООО «Газпромнефть НТЦ»

² Москва, Московский физико-технический институт

Особенности распространения осадочных пород в пространстве значительно определяют фильтрационно-емкостные свойства залежи, от которых в свою очередь зависят продуктивность и, в конечном счете, накопленная добыча скважин. Целью данной работы является построение метода для распознавания литолого-фациальной обстановки нефтяного месторождения на основе промысловых данных каротажных диаграмм.

В работе реализуется спектральный метод представления геофизических полей, который позволяет адекватно моделировать низкопроницаемые и высокорасчлененные коллектора. В качестве входных данных задачи рассматривается набор проинтерпретированных геологом каротажных данных, которые являются обучающей выборкой для алгоритма. Для сокращения размерности задачи осуществляется выбор информативных признаков (коэффициентов Фурье) на этапе обучения алгоритма. Далее методами машинного обучения решается задача классификации по типам фациальных обстановок каротажных данных со всего месторождения с использованием спектрального представления геофизических полей. В настоящей работе содержится также пример успешного применения предложенного метода для классификации фаций на участке реального месторождения [1].

- [1] *Белозеров Б. В., Бочков А. С., Урмаев О. С., Фукс О. М.* Методы машинного обучения при восстановлении обстановки осадконакопления в задачах геостатистического моделирования // Машинное обучение и анализ данных, 2014 (в печати).

Machine learning for sedimentation environment study in the tasks of geostatistics simulation

*Belozеров Борис*¹

Belozеров.BV@gazpromneft-ntc.ru

*Bochkov Andrey*¹

Bochkov.AS@gazpromneft-ntc.ru

*Ushmaev Oleg*¹

Ushmaev.OS@gazprom-neft.ru

Fuks Olga^{1,2*}

Phuks.OM@gazprom-neft.ru

¹Saint-Petersburg, LLC “Gazpromneft NTC”

²Moscow, Moscow Institute of Physics and Technology

Distribution properties of sedimentary rocks determine considerably the geometry and size of the reservoir and, consequently, the volume of hydrocarbon reserves. Therefore, knowledge about general patterns of rock sedimentation is significant for effective reservoir development. This work suggests the method to study the geological structure of oil field by automated recognition of litho-facial environment using geophysical field data.

In geophysical field representation, the spectral method is used which is well known for its adequate simulation of low-permeability and compartmentalized reservoirs. The input data are presented by geophysical information interpreted by the field geologist, which then form the training set for the machine-learning algorithm. To reduce the dimensionality of the data, only their significant features (Fourier coefficients) are retained in the learning step of the algorithm. Further, the data are classified into the different facial regions using the machine learning technique.

The described method was applied to the region of the real oil field (West Siberia) and allowed to successfully classify the wells according to the litho-facial sedimentation environment [1].

- [1] Belozеров B., Bochkov A., Ushmaev O., Fuks O. 2014 (in press). Machine learning for sedimentation environment study in the tasks of the geostatistics simulation. *Machine Learning and Data Analysis*.

Апостериорные алгоритмы в проблеме геофизического мониторинга

*Воскобойникова Гюльнара Маратовна*¹ gulya@opg.sscs.ru

*Кельманов Александр Васильевич*² kelm@math.nsc.ru

Хайретдинов Марат Саматович^{1,3*} marat@opg.sscs.ru

¹Новосибирск, Институт вычислительной математики и математической геофизики СО РАН

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

³Новосибирск, Новосибирский государственный технический университет

Предложен и исследуется новый подход в связи с решением проблемы активного геофизического мониторинга сейсмических процессов, порождаемых техногенными источниками. В основе мониторинга лежит решение задачи обнаружения и выделения геоакустических волновых форм с повышенной точностью. Задача оценивания неизвестных параметров источника сводится к решению нелинейной системы уравнений: $\hat{\eta} = \eta(\gamma, \chi) + \varepsilon$, где $\hat{\eta} = (\hat{n}_1, \dots, \hat{n}_N)^T$ — вектор измеренных времен пробега сейсмических волн; $\eta(\gamma, \chi) = (n_1, \dots, n_N)^T$ — N -мерный вектор вычисляемых времен пробега (теоретический годограф) или функция регрессии; $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)^T$ — вектор невязок; $\chi = (x, y, z, v, t)^T$ — m -мерный вектор оцениваемых параметров; $\gamma = (\gamma_1, \dots, \gamma_N)$ — матрица координат датчиков; N — число датчиков. Предлагается подход к совместному решению задачи измерения вектора $\hat{\eta} = (\hat{n}_1, \dots, \hat{n}_N)^T$ и исходной системы уравнений в едином процессе дискретной оптимизации. Решение достигается на основе применения апостериорных алгоритмов [1] к обнаружению и измерению параметров квазипериодических последовательностей сейсмических волновых форм. Эффективность подхода доказывается высокоточными результатами численного моделирования и экспериментов по мониторингу положения перемещающегося взрывного источника в глубокой скважине. Последние выполнены в связи с проблемой нефтепромыслового бурения.

- [1] *Voskoboinikova G. M., Khairatdinov M. S., Omelchenko O. K., Sedukhina G. F.* Numerical algorithms and results of experiments to determine the parameters of the borehole bottom and medium // Bulletin of the Novosibirsk Computing Center, 2005. Iss. 10. P. 35–43.

Posteriori algorithms in a problem of geophysical monitoring

*Voskoboynikova Gyl'nara*¹

gulya@opg.sbcc.ru

*Kel'manov Alexander*²

kelm@math.nsc.ru

Khairtdinov Marat^{1,3}✉

marat@opg.sbcc.ru

¹Novosibirsk, Institute Computational Mathematics and Mathematical Geophysics SB RAS

²Novosibirsk, Sobolev Institute of Mathematics SB RAS

³Novosibirsk, Novosibirsk, State Technical University

The new approach to the problem of active geophysical monitoring of the seismic processes generated by technogenic sources is suggested and investigated. Monitoring is based on the problem of detection and selection of geoacoustic wave forms with higher accuracy. The problem of estimation of unknown parameters of a source is led to the solution of the nonlinear equations system: $\hat{\eta} = \eta(\gamma, \chi) + \varepsilon$ where $\hat{\eta} = (\hat{\eta}_1, \dots, \hat{\eta}_N)^T$ is the vector of the measured arrival times of seismic waves; $\eta(\gamma, \chi) = (n_1, \dots, n_N)^T$ is the N -dimensional vector of a theoretical arrival times or the regression function; $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)^T$ is the residual vector; $\chi = (x, y, z, v, t)^T$ is the m -dimensional vector of the estimated parameters; $\gamma = (\gamma_1, \dots, \gamma_N)$ is the matrix of the sensors coordinates; and N is the number of sensors.

The approach to the joint solution of the vector measurement $\hat{\eta} = (\hat{\eta}_1, \dots, \hat{\eta}_N)^T$ and original system equation is suggested in uniform process of discrete optimization. The solution is reached with application of posteriori algorithms [1] for detection and measurement of quasi-periodic sequences parameters of seismic wave forms. Efficiency of the approach is proved by high-precision results of numerical modeling and experiments on monitoring of moving explosion source in a deep borehole. They are executed for the problem of oil-field drilling.

- [1] Voskoboynikova G. M., Khairtdinov M. S., Omelchenko O. K., Sedukhina G. F. 2005. Numerical algorithms and results of experiments to determine the parameters of the borehole bottom and medium. *Bulletin of the Novosibirsk Computing Center* 10:35–43.

Научная информационная система «Активная сейсмология» для комплексных геофизических исследований

*Ковалевский Валерий Викторович** kovalovsky@sscc.ru
Брагинская Людмила Петровна ludmila@opg.sccc.ru
Григорюк Андрей Павлович and@opg.sccc.ru
Кратов Сергей Викторович kratov@sscc.ru

Новосибирск, Институт вычислительной математики
и математической геофизики СО РАН

Представлена web-ориентированная Научная информационная система (НИС) «Активная сейсмология» для комплексного информационного обеспечения теоретических и экспериментальных исследований в новом направлении геофизики — изучении строения земной коры и исследования геодинамических процессов в зонах землетрясений и вулканов с использованием управляемых источников сейсмических волн. Основной задачей НИС является получение, интеграция и предоставление данных и знаний в интересах фундаментальных научных исследований физики сейсмического процесса, изучения глубинного строения геологической среды, мониторинга природных сейсмических и вулканических процессов. Одна из возможностей НИС — применение методов обработки и интерпретации многопараметрических геофизических данных с использованием новых информационных технологий. Система позволяет пользователю проводить в удаленном доступе исследовательские работы с данными уникального архива полевых экспериментов 1995–2014 гг., включающего файлы волновых форм и сопутствующей информации. Научная информационная система «Активная сейсмология» охватывает все основные этапы научных исследований в активной сейсмологии: эксперимент, моделирование, библиографию, публикацию результатов и их обсуждение [1].

Работа поддержана грантами РФФИ № 14-07-00832 и № 12-05-00786.

[1] <http://opg.sccc.ru>

Scientific Information System “Active Seismology” for integrated geophysical studies

*Kovalevskiy Valeriy**

kovalevsky@sscc.ru

Braginskaya Ludmila

ludmila@opg.sccc.ru

Grigoruk Andrey

and@opg.sccc.ru

Kratov Sergey

kratov@sscc.ru

Novosibirsk, Institute of Computational Mathematics and Mathematical
Geophysics SB RAS

The paper presents a web-oriented Scientific Information System (SIS) “Active Seismology” to provide comprehensive information for theoretical and experimental research in a new field of Geophysics — for studying of the crustal structure and geodynamic processes in the areas of earthquakes and volcanoes using managed sources of seismic waves. The main objective of SIS is receiving, integration, and provision of data and knowledge for fundamental physics research of seismic process, studying the deep structure of the geological environment, monitoring natural seismic and volcanic processes. One of the possibilities of SIS is the application for processing and interpretation of multiparameter geophysical data using new information technologies. The system allows user to perform remote access to research with unique archive data field experiments in 1995–2014, including waveform files and related information. Scientific Information System “Active Seismology” covers all major stages of research in active seismology: experiment, modeling, bibliography, publication of results, and their discussion [1].

This research is funded by RFBR, grants 14-07-00832 and 12-05-00786.

[1] <http://opg.sccc.ru>

Оценка состояния геомагнитного поля на основе совмещения вейвлет-преобразования с радиальными нейронными сетями

*Мандрикова Оксана Викторовна*¹ oksanam1@mail.ru
Жижикина Екатерина Андреевна^{2*} ekaterinazh1@mail.ru

¹с. Паратунка, Камчатский край, Институт космических исследований и распространения радиоволн ДВО РАН

²Петропавловск-Камчатский, Камчатский государственный технический университет

Работа направлена на создание теоретических и программных средств анализа параметров геомагнитного поля и оценки его состояния на основе современных методов распознавания образов и цифровой обработки сигналов. Существующие методы анализа геомагнитных данных не позволяют выявлять некоторые закономерности в данных и приводят к потере важной информации.

Предлагается метод, основанный на совместном применении вейвлет-преобразования с радиальными нейронными сетями. Метод позволяет изучать тонкие особенности структуры геомагнитных данных и выделять информативные составляющие, характеризующие степень возмущенности геомагнитного поля.

В работе детально изучена структура геомагнитных данных, определены признаки возрастания геомагнитной активности, предложены классы и способ формирования радиального слоя нейронной сети, позволяющий уменьшить количество используемых примеров и повысить качество классификации геомагнитных данных. На основе комбинации решений построенных программных нейросетевых систем предложено решающее правило по оценке состояния геомагнитного поля в автоматическом режиме. Апробация метода, выполненная на данных станции «Паратунка» (Камчатский край), подтвердила его эффективность [1].

Работа поддержана грантом РНФ № 14-11-00194.

- [1] *Мандрикова О. В., Жижикина Е. А.* Оценка состояния геомагнитного поля на основе совмещения вейвлет-преобразования с радиальными нейронными сетями // Машинное обучение и анализ данных, 2014 (в печати).

Estimation of degree of the geomagnetic field disturbance based on the combined use of wavelet transform with radial neural networks

Mandrikova Oksana¹

oksanam1mail.ru

Zhizhikina Ekaterina²*

ekaterinazh1@mail.ru

¹Paratunka, Kamchatka Region, Institute of Cosmophysical Research and Radio Wave Propagation

²Petropavlovsk-Kamchatsky, Kamchatka State Technical University

The present paper is focused on the development of theoretical tools and software for the analysis of the geomagnetic field parameters and for the estimation of its condition using modern methods of pattern recognition and digital signal processing. Existing methods for the geomagnetic data analysis do not allow to identify some regularities in the data and lead to loss of important information.

A method based on the combined use of the wavelet transform and radial neural networks has been proposed. This method allows one to study subtle structural features of the geomagnetic data and to extract informative components which characterize the disturbance degree of the geomagnetic field.

In the present paper, geomagnetic data structure was studied in detail, the signs of the geomagnetic activity increasing were defined, and classes for the radial layer of the neural network were suggested. Furthermore, a way of forming a radial layer was proposed. This way allows to significantly reduce the number of examples and to improve the quality of the geomagnetic data classification. On the basis of combination of decisions of the developed neural networks, a decision rule is proposed to estimate the geomagnetic field condition in the automatic mode. The method has been successfully tested on the geomagnetic data obtained from the Paratunka observatory (Kamchatka region) [1].

This research is funded by Russian Science Foundation, grant 14-11-00194.

- [1] Mandrikova O., Zhizhikina E. 2014 (in press). Estimation of degree of the geomagnetic field disturbance based on the combined use of wavelet transform with radial neural networks. *Machine Learning and Data Analysis*.

Моделирование вариаций космических лучей и выделение аномалий на основе совмещения вейвлет-преобразования с нейронными сетями

Мандрикова Оксана Викторовна^{1*} oksanam1@mail.ru
Залев Тимур Ленарович^{1,2} tim.aka.geralt@mail.ru

¹Паратунка, Камчатский край, Институт космофизических исследований и распространения радиоволн ДВО РАН

²г. Петропавловск-Камчатский, Камчатский государственный технический университет

Исследованы данные нейтронных мониторов станций «Афины», «Новосибирск» и «Апатиты» за 2005–2013 гг. и обнаружены аномальные особенности, возникающие в вариациях космических лучей во время сильных магнитных бурь. Исследования основаны на разработанном авторами методе моделирования компонент данных космических лучей путем совмещения вейвлет-преобразования и нейронных сетей прямого распространения. Выполняется кратномасштабное вейвлет-разложение данных и выделяются информативные компоненты. Полученные компоненты аппроксимируются нейронными сетями прямого распространения. Метод позволяет выполнить детальный анализ структуры данных и путем анализа ошибок нейронной сети выявить аномальные особенности (Форбуш-эффекты) во временном ходе космических лучей [1].

Работа выполнена при поддержке гранта РНФ №14-11-00194.

- [1] *Мандрикова О. В., Залев Т. Л.* Моделирование вариаций космических лучей и выделение аномалий на основе совмещения вейвлет-преобразования с нейронными сетями // Машинное обучение и анализ данных, 2014 (в печати).

Modeling of cosmic ray variations and allocation of anomalies based on a combination of wavelet transform with neural networks

*Mandrikova Oksana*¹★

oksanam1@mail.ru

Zalyaev Timur^{1,2}

tim.aka.geralt@mail.ru

¹Paratunka, Kamchatka Region, Institute of Cosmophysical Researches and Radio Wave Propagation of the Far Eastern Branch of the Russian Academy of Sciences

²Petropavlovsk-Kamchatsky, Kamchatka State Technical University

Valuable information about the topology change of the geomagnetic field during magnetic storms is provided by study of the dynamics of cosmic rays. In disturbed periods, the recorded parameters of the cosmic rays have a complex nonstationary structure, contain nonsmooth local features, which occur at random time moments and carry important information about the studied processes. Lack of theoretical apparatus providing an adequate description of the analyzed data leads to an inevitable loss and distortion of the information and requires advanced methods, among which are of great importance the methods of pattern recognition and digital signal processing.

Based on a combination of multiresolution wavelet decompositions with neural networks, a method of approximation of the cosmic rays time course and the allocation of anomalous variations (Forbush effects) that occur during the periods of high solar activity has been proposed.

On the basis of the proposed method for the stations “Novosibirsk,” “Apatity,” and “Athens,” software systems were built for neural network approximation of typical variations of cosmic rays and the analysis of data in the periods of strong magnetic storms.

Application of the method in conjunction with other methods and approaches allows better perform the assessment of the state of space weather [1].

This research is funded by Russian Science Foundation, grant 14-11-00194.

- [1] Mandrikova O., Zalyaev T. 2014 (in press). Modeling of of cosmic ray variations and allocation of anomalies based on a combination of wavelet transform with neural networks. *Machine Learning and Data Analysis*.

Сжатие гиперспектральных данных на основе метода кодирования с преобразованием

Чичева Марина Александровна^{1*}

mchi@geosamara.ru

Юзькив Руслан Романович^{1,2}

yuzkiv@rambler.ru

¹ Самара, Самарский государственный аэрокосмический университет имени академика С. П. Королёва (национальный исследовательский университет)

² Самара, Институт систем обработки изображений РАН

В ряду задач обработки данных особое место занимают задачи, связанные с дистанционным зондированием Земли из космоса. В последнее время наибольший интерес вызывают так называемые гиперспектральные данные (или изображения), представляющие собой трехмерный массив, состоящий из ряда изображений одного и того же участка земной поверхности, полученных в разных спектральных диапазонах.

Как правило, каждый спектральный слой в этой структуре содержит двухбайтные данные, т. е. целые числа в диапазоне от 0 до 65 535. Очевидно, что объем каждого гиперспектрального изображения весьма велик, что приводит к сложностям с хранением и передачей таких данных. В этой связи особенно актуальной становится задача их сжатия.

В настоящей работе рассматривается обобщение известного метода кодирования с преобразованием на случай гиперспектральных данных. Предложенный подход к сжатию и блочному хранению сжатых данных позволяет восстанавливать из архива как все данные целиком, так и отдельные пространственные срезы или кривые спектрального отражения в заданной точке.

Проведено исследование работы алгоритма в различных режимах. Выработаны рекомендации, обоснован выбор отдельных составляющих метода. Выполнено сравнение с известными методами сжатия гиперспектральных данных, показано преимущество по сравнению с рядом алгоритмов [1].

Работа поддержана грантами РФФИ №№ 12-07-00751, 12-01-00822 и 13-07-97006.

[1] *Чичева М. А., Юзькив Р. Р.* Сжатие гиперспектральных данных на основе метода кодирования с преобразованием // Компьютерная оптика, 2014 (в печати). Т. 3.

Hyperspectral data compression based on transform coding method

Chicheva Marina^{1*}

mchi@geosamara.ru

Yuzkiv Ruslan^{1,2}

yuzkiv@rambler.ru

¹Samara, Samara State Aerospace University

²Samara, Image Processing System Institute of RAS

The problems dealing with Earth's remote sensing have a special place in the list of data processing tasks. At the present time, the main interest as drawn to so-called hyperspectral data (or images), which are three-dimensional array consisting of image series for the same Earth surface field in different spectral bands.

As a rule, each spectral layer in this structure contains double-byte data, i. e., integers in the band from 0 to 65 535. It is obviously that the volume of any hyperspectral image is extra large; so, its storage and transfer are complex. Therefore, the problem of compression is extremely actual for such data.

The generalization of the known transform coding method on the case of hyperspectral data is considered.

Proposed approach to compression and blocked storage of compressed data allows one to reconstruct from an archive both all data as a whole and certain spatial face or spectral-reflection curve at the special point.

The analysis of algorithm implementation in the different modes has been made. The recommendations have been produced and the choice of some steps has been justified. The comparison with known methods of hyperspectral data compression is done, the advantage in comparison with some approaches is shown [1].

This research is funded by RFBR, grants 12-07-00751, 12-01-00822, and 13-07-97006.

[1] Chicheva M., Yuzkiv R. 2014 (in press). Hyperspectral data compression based on transform coding method. *Computer Optics* 3.

Интеллектуальные методы обнаружения и классификации сетевых атак

*Браницкий Александр Александрович** alexander.branitskiy@gmail.com

Тимофеев Адиль Васильевич

tav@iias.spb.su

Санкт-Петербург, СПИИРАН

Задача обнаружения сетевых атак является очень актуальной: глобальное развитие компьютерных сетей и многочисленные хакерские атаки порождают множество проблем, связанных с безопасностью информационных ресурсов. Несовершенство существующих методов защиты информации, а также изменяющийся характер злонамеренных действий со стороны атакующего приводят систему в небезопасное состояние. Поэтому становится важно идентифицировать атаки и своевременно реагировать на них. В статье представлены адаптивные подходы к процессу обнаружения злоупотреблений, которые используют аналитические возможности нейросетей, иммунных клеток и интерполяционных полиномов, и показаны результаты анализа каждого метода. Основное преимущество предложенных подходов заключается в обнаружении ранее неизвестных типов атак. Для распознавания образцов соединений применяются три математические модели:

- система из многослойных нейронных сетей, статически обученных по правилу обратного распространения ошибки;
- система из иммунных клеток, представляющих собой самоорганизующиеся карты Кохонена и динамически обучающихся по конкурентному алгоритму;
- система из многомерных интерполяционных полиномов, описывающих пространство признаков элементов контрольного множества.

На их основе разработано программное средство, предназначенное для анализа сетевого трафика на наличие аномальных соединений, а также для их классификации при обмене данными по компьютерным сетям, использующим стек протоколов ТСП/IP [1].

- [1] *Браницкий А. А., Тимофеев А. В.* Интеллектуальные методы обнаружения и классификации сетевых атак // Машинное обучение и анализ данных, 2014 (в печати).

Intellectual methods of detection and classification of network attacks

*Branitskiy Alexander**

alexander.branitskiy@gmail.com

Timofeev Adil

tav@iias.spb.su

Saint-Petersburg, SPIIRAS

The task of detection of network attacks is very relevant: global development of computer networks and numerous hacker attacks generate a lot of problems related with security of information resources. Imperfection of the existing methods of information security as well as the changing nature of malicious actions of the attacker lead to unsafe state of the system. Therefore, it is important to identify attacks and timely respond to their emergence. The paper describes the adaptive approaches to the process of a misuse detection, which use analytical capabilities of neural networks, immune cells, and interpolation polynomials; also, the results of analysis of each method are demonstrated. The main advantage of the proposed approaches is detection of previously unknown attacks. For recognition of instances of network connections, three mathematical models are used:

- a system of multilayer neural networks statically trained by the algorithm of error backpropagation;
- a system of immune cells which are self-organizing Kohonen's maps and dynamically training by the competitive algorithm; and
- a system of multivariate interpolation polynomials describing the feature space of elements of the control set.

A software tool was developed on their basis to analyze network traffic for anomalous connections and to classify them during the transmission of data over the computer networks using a protocol stack TCP/IP [1].

- [1] Branitskiy A., Timofeev A. 2014 (in press). Intellectual methods of detection and classification of network attacks. *Machine Learning and Data Analysis*.

Разработка прототипа централизованной системы избирательной многофакторной биометрической аутентификации

*Ручай Алексей Николаевич**

ruchai@pochta.ru

Горшенин Владимир Викторович

62612740@mail.ru

Маткин Илья Александрович

ilya.matkin@gmail.com

Челябинск, Челябинский государственный университет

Современные тенденции показывают стремление использовать системный подход для создания мультибиометрических систем аутентификации, так как существуют программные реализации на основе, как правило, только одной биометрической характеристики без дополнительных инструментов и модулей, что создает проблемы при их использовании и эксплуатации.

В [1] был описан разработанный комплекс модулей биометрической аутентификации для разграничения прав доступа в ОС Windows XP. Данный комплекс модулей послужил основой для разработки централизованной системы разграничения прав доступа на основе избирательной многофакторной биометрической аутентификации с клиент-серверной архитектурой. Для этого были разработаны принципы построения системы, архитектуры, протокола взаимодействия и был реализован центр и модули биометрической аутентификации.

В зависимости от разных условий и факторов, в частности от доступности электронных средств, от удобства, от стойкости к атакам и уязвимостям, от болезней или увечий пользователей, может быть выбрана биометрическая аутентификация на основе любых таких биометрических характеристик, как ритм ввода пароля, голос, динамика подписи и графическое распознавание.

Важным этапом была разработка архитектуры системы и протокола передачи биометрических данных. Кроме того, был проведен анализ возможных атак на разработанную систему и сделаны выводы и рекомендации по методам защиты от них.

Работа поддержана грантом РФФИ № 14-07-31049.

- [1] *Ручай А. Н.* Разработка универсального комплекса модулей для разграничения прав доступа в ОС Windows XP на основе биометрической аутентификации // *Безопасность информационных технологий*, 2013. С. 74–78.

Development of prototype centralized system of election multifactor biometric authentication

*Ruchay Alexey**

ruchai@pochta.ru

Gorshenin Vladimir

gorshenin.vladimir@gmail.com

Matkin Ilya

ilya.matkin@gmail.com

Chelyabinsk, Chelyabinsk State University

Current trends show a desire to use a systematic approach for creating multibiometric authentication systems as there are software implementation based, as a rule, on only one biometric feature without additional tools and modules that creates problems in their applying and maintenance.

In [1], the developed complex of modules of biometric authentication for managing permissions on Windows XP OS was described. This complex modules formed a basis for development of the centralized system of managing permissions based on the election multibiometric authentication with client-server architecture. For this purpose, the principles of system creation, architecture, and communication protocols were developed and also, the center and modules of biometric authentication was realized.

Depending on various conditions and factors, including the availability of electronic means and convenience, resistance to attacks and exploits, disease or injury of a user biometric authentication can be selected on the basis of any biometrics such as rhythm password, voice, dynamic signatures, and graphics recognition.

Development of the system architecture and biometric data communication protocol was an important stage. Besides, the possible attacks on the developed system were analyzed. The conclusions and recommendations on defenses from these attacks are submitted.

This research is funded by RFBR, grant 14-07-31049.

- [1] Ruchay A. 2013. Development of a universal set of modules for managing permissions Windows XP on based biometric authentication. *Security of Information Technology*. 74–78.

Идентификация потенциально опасных клиентских запросов на основе многоклассового распознавания образов

*Сычугов Алексей Алексеевич**

xru2003@list.ru

Никишин Михаил Сергеевич

j48sa150a.p3az@gmail.com

Середин Олег Сергеевич

oseredin@yandex.ru

Тула, Тульский государственный университет

В настоящее время, с ростом количества клиентских запросов к серверам и объема обрабатываемого трафика, а также с возрастающим объемом потенциально опасных клиентских запросов, все актуальнее становится задача идентификации и выявления таких запросов с целью предупреждения возникновения той или иной сетевой атаки. Имеющиеся в настоящее время системы обнаружения вторжений (СОВ) эффективно решают эту задачу, однако постоянно увеличивающийся поток запросов, появление как неизвестных вторжений, так и новых реализаций известных сетевых атак заставляет совершенствовать СОВ, а именно те из них, которые направлены на обнаружение аномалий в сетевом взаимодействии.

Таким образом, задача разработки новых методов обнаружения вторжений не теряет своей актуальности. В данной работе для идентификации потенциально опасных клиентских запросов предлагается использовать метод многоклассового распознавания образов, основанный на методе двухклассовой классификации с нечетким решающим правилом. Выбор данного метода обусловлен тем, что идентификация клиентских запросов фактически является задачей многоклассовой идентификации на больших массивах данных, которая является ресурсоемкой, а предлагаемый метод позволяет снизить вычислительную сложность, тем самым повысить эффективность идентификации [1].

Работа поддержана грантом РФФИ № 14-07-00527-а.

- [1] Двоенко С. Д., Моттль В. В., Середин О. С. Процедура распознавания образов для случая многих классов, опирающаяся на совокупность функций степени достоверности для пар классов // Известия ТулГУ, Вычислительная техника, автоматика, управление, 1999. Т. 2. Вып. 2. С. 28–35.

Detection of potentially dangerous client inquiries on the basis of multiclass pattern recognition

*Sychugov Alexey**

xru2003@list.ru

Nikishin Michael

j48sa150a.p3az@gmail.com

Seredin Oleg

oseredin@yandex.ru

Tula, Tula State University

The growing number of client inquiries to servers and the increasing intensity of traffic to be processed as well as the growing number of potentially dangerous inquiries drastically actualize the problem of detection and identification of such inquiries with the purpose of preventing network attacks. The existing network invasion detection systems effectively solve this problem, but the increasing flow of inquiries coupled with the emergence of new unknown invasions and new realizations of known network attacks strongly demand further enhancement of invasion detection systems, primarily, those of them that are aimed at detecting anomalies in network interaction.

Thus, the necessity of developing new methods of invasion detection does not lose its actuality. In this work, a new approach to detecting potentially dangerous client inquiries is proposed on the basis of multiclass pattern recognition, which is underlied, in its turn, by the method of fuzzy dichotomy. This method has been chosen because the client inquiry identification is, actually, a problem of multiclass pattern recognition in large data sets, whereas combining several two-class classifiers promises to essentially decrease the computational complexity of multiclass pattern recognition and, thereby, to increase the effectiveness of intrusion detection [1].

This research is funded by RFBR, grant 14-07-00661-a.

- [1] Dvoenko S.D., Mottl V.V., Seredin O.S. 1999. A pattern recognition procedure for the case of several classes on the basis of the set of confidence functions for pairs of classes. *Izvestiya of Tula State University, Computers, Automation and Control* 2(2):28–35. [In Russian.]

Содержание

Теория классификации	8
<i>Абрамов В. И., Середин О. С., Моттль В. В.</i>	
Обучение распознаванию образов в произвольных метрических пространствах	8
<i>Журавлев Ю. И.</i>	
Об одной модификации алгоритмов вычисления оценок	10
<i>Загоруйко Н. Г., Кутненко О. А., Зырянов А. О., Леванов Д. А.</i>	
Обучение распознаванию без переобучения	12
<i>Левдик П. В., Черноусова Е. О., Сунгуров Д. С., Моттль В. В.</i>	
Экспериментальное исследование метода неявной кросс-валидации в задаче беспризнакового обучения распознаванию образов	14
<i>Лепский А. Е.</i>	
Сравнение искаженных гистограмм вероятностными методами	16
<i>Махина Г. А.</i>	
Обучение монотонного классификатора ближайшего соседа по немонотонной выборке	18
<i>Неделько В. М.</i>	
Некоторые вопросы оценивания качества методов построения решающих функций	20
<i>Никишин М. С., Красоткина О. В., Моттль В. В.</i>	
Обучение многоклассовому распознаванию образов по большим обучающим совокупностям	22
<i>Стенина М. М., Кузнецов М. П.</i>	
Комбинирование частичных порядков в задаче обучения по предпочтениям	24
<i>Татарчук А. И., Моттль В. В.</i>	
Байесовские постановки задачи обучения распознаванию образов по методу опорных векторов с управляемой селективностью отбора информативных признаков объектов	26

<i>Турков П. А., Красоткина О. В.</i>	
Отбор признаков в задаче обучения распознаванию образов при смещении решающего правила	28
<i>Черноусова Е. О., Левдик П. В., Моттль В. В.</i>	
Неявная кросс-валидация для выбора подмножества информативных признаков в задаче обучения распознаванию образов по методу опорных векторов	30
<i>Чуличков А. И., Юань Б.</i>	
Оценки, минимизирующие возможность потерь, и минимаксные оценки: сравнительный анализ	32
Метрические методы анализа данных	34
<i>Бериков В. Б.</i>	
Ансамбль алгоритмов кластерного анализа с весами	34
<i>Двоенко С. Д.</i>	
Двухкомпонентная функция качества кластеризации множества элементов, представленных парными сравнениями	36
<i>Двоенко С. Д., Пшеничный Д. О.</i>	
Оптимальная коррекция метрических нарушений в матрицах парных сравнений	38
<i>Иофина Г. В., Минаев А. В., Поляков Ю. С., Максимов Ю. В.</i>	
Выбор метрики в задачах частичного обучения	40
<i>Пушняков А. С.</i>	
О комбинаторных оценках максимальных ε -разбиений метрических конфигураций	42
<i>Рудаков К. В.</i>	
О реляционных факторизациях полуметрического конуса	44
Дискретно-логические методы классификации	46
<i>Генрихов И. Е.</i>	
О критериях ветвления, используемых при синтезе решающих деревьев	46
<i>Дюкова Е. В., Никифоров А. Г., Прокофьев П. А.</i>	
Эффективный параллельный алгоритм дуализации	48

<i>Дюкова Е. В., Прокофьев П. А.</i>	
Новые асимптотически оптимальные алгоритмы дуализации	50
<i>Покровская И. В., Гольдовская М. Д., Дорофеев Ю. А., Киселёва Н. Е.</i>	
Методы интеллектуальной обработки качественных данных	52
<i>Торшин И. Ю., Рудаков К. В.</i>	
Инварианты размеченных графов в комбинаторной теории разрешимости	54
<i>Янковская А. Е., Ямианов А. В., Кривдюк Н. М.</i>	
Когнитивные средства в интеллектуальных системах поддержки принятия решений	56
Анализ временных рядов	58
<i>Загоруйко Н. Г., Борисова И. А., Леванов Д. А.</i>	
Описание многомерных динамических процессов на языке иерархии концептов	58
<i>Загоруйко Н. Г., Татарников В. В.</i>	
Использование избыточности в кубах данных для обнаружения ошибок и заполнения пробелов	60
<i>Теклина Л. Г., Котельников И. В.</i>	
Синтез линейных многомерных систем квазиинвариантного управления методами распознавания образов	62
<i>Филиппенков Н. В., Петрова М. А.</i>	
О некоторых вопросах анализа пучков временных рядов	64
<i>Чернявский А. Л., Дорофеев А. А., Дорофеев Ю. А.</i>	
Интеллектуальные методы структурного анализа многомерных траекторий	66
<i>Янковская А. Е., Аметов Р. В., Китлер С. В.</i>	
К вопросу о создании самообучающейся системы поддержки принятия решений в области биржевой торговли, основанной на отказоустойчивых подмножествах сигнальных признаков	68

Сложность вычислений и оптимизация	70
<i>Агеев А. А., Кельманов А. В., Пяткин А. В.</i>	
NP-трудность евклидовой задачи о максимальном разрезе	70
<i>Борисов Б. Д.</i>	
Оптимальная оценка базовой характеристики нестабильности частоты осцилляторов	72
<i>Волков Ю. С., Мирошниченко В. Л., Салиенко А. Е.</i>	
Математическое моделирование универсальной характеристики поворотно-лопастной гидротурбины	74
<i>Гасников А. В., Крымова Е. А.</i>	
Стохастическая онлайн оптимизация в случае тяжелых хвостов стохастических градиентов	76
<i>Гимади Э. Х., Истомин А. М., Рыков И. А., Цицулко О. Ю.</i>	
Об асимптотической разрешимости задачи t коммивояжеров на случайных входах, неограниченных сверху	78
<i>Гимади Э. Х., Кельманов А. В., Пяткин А. В., Хачай М. Ю.</i>	
Приближенные алгоритмы поиска в графе несмежных клик минимального суммарного веса	80
<i>Глинский Б. М., Марченко М. А., Родионов А. С., Каравачев Д. А., Поджорытов Д. И.</i>	
Отображения параллельных алгоритмов на суперкомпьютеры экзафлопсной производительности на основе имитационного моделирования	82
<i>Горнов А. Ю., Зароднюк Т. С.</i>	
Вычислительная технология оценки степени выпуклости многоэкстремальной функции	84
<i>Иофина Г. В., Максимов Ю. В.</i>	
Вычислительная сложность минимизации булевых функций в классе схем глубины 2	86
<i>Карацуба Е. А.</i>	
Быстрые алгоритмы. Представление проекта.	88
<i>Кельманов А. В.</i>	
Полиномиальные алгоритмы с гарантированными оценками точности для некоторых задач анализа и распознавания последовательностей	90

<i>Кельманов А. В., Хамидуллин С. А.</i>	
Приближенный полиномиальный алгоритм для одной задачи разбиения последовательности	92
<i>Кельманов А. В., Хандеев В. И.</i>	
Точный псевдополиномиальный алгоритм для одной задачи двухкластерного разбиения множества векторов . .	94
<i>Кириш Д. В., Куприянов А. В.</i>	
Идентификация кристаллических решеток на основе оценивания параметров элементарных ячеек Браве и Вигнера–Зейтца	96
<i>Хачай М. Ю., Незнакина Е. Д.</i>	
Аппроксимируемость геометрической задачи о нескольких коммивояжерах	98
<i>Чернов В. М.</i>	
Тернарная машинная арифметика: история, проблемы, приложения	100
<i>Чихрадзе К. К., Коршунов А. В., Кузюрин Н. Н.</i>	
Использование модели социальной сети с сообществами пользователей для распределенной генерации случайных социальных графов	102
Обработка изображений	104
<i>Грачева И. А., Копылов А. В., Красоткина О. В.</i>	
Адаптивная процедура обобщенного сглаживания изображений на основе статистического подхода	104
<i>Каржищенко А. Н., Мнухин В. Б.</i>	
Симметризация точек изображения, заданных статистическими выборками	106
<i>Ларин А. О.</i>	
Разработка трехмерного сканера человеческого тела для проекта виртуальной примерочной с использованием сенсоров Kinect	108
<i>Федотов Н. Г., Сёмов А. А., Моисеев А. В.</i>	
Интеллектуальные возможности гипертрейс-преобразования: конструирование признаков с заданными свойствами . .	110

Анализ изображений	112
<i>Егоров А. И., Вшивняков Б. В.</i>	
Метод полуавтоматической контекстной разметки сцены	112
<i>Мурашов Д. М., Березин А. В., Иванова Е. Ю.</i>	
Сравнение изображений картин по информативным фрагментам	114
<i>Кий К. И.</i>	
Новый метод реального времени для контекстного описа- ния изображений и его применения к навигации роботов и интеллектуальному управлению	116
<i>Корнилов Ф. А.</i>	
Поиск структурных различий изображений	118
<i>Каржищенко А. Н., Мнухин В. Б.</i>	
Топологическая фильтрация для распознавания и анали- за симметрии цифровых изображений	120
Анализ видеопоследовательностей	122
<i>Каратеев С. Л., Костромов Н. А., Бекетова И. В., Визиль- тер Ю. В.</i>	
Комплексирование информации с использованием N- грамм при распознавании символьных последовательно- стей на видеопоследовательностях	122
<i>Карпов Р. О., Копылов А. В.</i>	
Использование кластерного анализа для цифровой ста- билизации видеопоследовательностей	124
<i>Лебедев М. А., Бондаренко М. А., Комаров Д. В., Степа- нъянц Д. Г., Выголов О. В., Визильтер Ю. В., Жел- тов С. Ю.</i>	
Алгоритм автоматического совмещения сенсорной и син- тезируемой видеоинформации для авиационной системы комбинированного видения	126
<i>Митекин В. А., Федосеев В. А.</i>	
Метод встраивания информации в видеопоследователь- ности, стойкий к ошибкам потери синхронизации	128

<i>Морозов А. А., Вайш А., Полупанов А. Ф., Анциперов В. Е., Лычков И. И., Алфимцев А. Н., Девятков В. В.</i>	
Разработка исследовательской программной платформы для параллельного объектно-ориентированного логического программирования интеллектуального видеонаблюдения	130
<i>Панов А. И.</i>	
Алгебраические свойства операторов распознавания в моделях зрительного восприятия динамических сцен . . .	132
<i>Ташлинский А. Г., Воронов С. В.</i>	
Использование взаимной информации как целевой функции качества оценивания параметров изображений	134
<i>Черноусов В. О., Савченко А. В.</i>	
О зависимости точности морфологического алгоритма обнаружения груза на вилочном погрузчике от интенсивности аддитивного шума	136
Распознавание изображений	138
<i>Бекетова И. В., Каратеев С. Л., Ососков М. В., Костромов Н. А.</i>	
Технология получения стандартного изображения лица с использованием программно-аппаратного комплекса подготовки и контроля цифровых фотографий и 3D-модели поверхности лица	138
<i>Кушнир О. А., Середин О. С.</i>	
Применение процедуры сравнения скелетов бинарных изображений к задаче распознавания типов лекарственных растений	140
<i>Ланге М. М., Ганебных С. Н.</i>	
Эффективность иерархических алгоритмов классификации двумерных объектов в терминах соотношения качество–сложность	142
<i>Матвеев И. А., Новик В. П.</i>	
Способы выбора наилучшего эталона радужки из последовательности	144

Анализ формы изображений	146
<i>Визильтер Ю. В., Желтов С. Ю.</i>	
Меры сходства и метрики сравнения форм мозаичных изображений	146
<i>Горбачевич В. С., Визильтер Ю. В., Желтов С. Ю., Рубис А. Ю., Воротников А. В.</i>	
Морфлеты: новый класс древовидных морфологических описаний формы изображений на основе систем хаароподобных вейвлетов	148
<i>Местецкий Л. М.</i>	
Медиальная ширина фигуры — дескриптор формы изображений	150
<i>Пытьев Ю. П.</i>	
Косые проекторы, относительные формы и субъективные модели в морфологии изображений	152
<i>Сидякин С. В., Визильтер Ю. В.</i>	
Морфологические дескрипторы формы бинарных изображений на основе эллиптических структурирующих элементов	154
<i>Степанов А. В., Середин О. С., Кушнир О. А.</i>	
Автоматический подбор адекватных значений параметров регуляризации, аппроксимации и склейки скелетных графов бинарных изображений	156
Приложения: медицина	158
<i>Анциперов В. Е.</i>	
Обнаружение ритмов головного мозга человека на основе величины корреляции аналитических спектров электроэнцефалограммы в основных диапазонах частот	158
<i>Берестнева О. Г., Воловоденко В. А., Гергет О. М., Шаропин К. А., Осадчая И. А.</i>	
Методы визуализации многомерных медицинских данных на основе обобщенных графических образов	160
<i>Макаров В. Л., Бекларян Л. А., Белоусов Ф. А.</i>	
Стационарные режимы в модели Хенинга и ее модификациях	162

<i>Оплачко Е. С., Рыжунев С. Д., Устинин М. Н.</i> Облачный ресурс MathBrain для анализа данных энцефалографии	164
<i>Панкратова Н. М., Устинин М. Н.</i> Спектральный анализ высокочастотной области данных магнитной энцефалографии	166
<i>Рыжунев С. Д., Сычев В. В., Устинин М. Н.</i> Выделение и локализация источников аудиторного отклика коры головного мозга по сильно зашумленным данным магнито- и электроэнцефалографии	168
<i>Сушкова О. С., Королев М. С., Габова А. В., Иллариошкин С. Н., Карabanов А. В., Обухов К. Ю., Обухов Ю. В.</i> Совместный анализ сигналов электроэнцефалограмм, электромиограмм и механического тремора при болезни Паркинсона в ранней стадии	170
<i>Успенский В. М., Воронцов К. В., Целых В. Р., Бунаков В. А., Ефимова И. В., Полежаев В. А.</i> Информационный анализ электрокардиосигналов для диагностики многих заболеваний внутренних органов по одной электрокардиограмме	172
<i>Устинин М. Н.</i> Функциональная томография мозга человека на основе анализа энцефалограмм в пространстве частота–паттерн	174
Приложения: биоинформатика	176
<i>Кузнецов Е. Н., Анашкина А. А., Есипова Н. Г., Туманян В. Г.</i> Кластер-анализ пространственных контактов аминокислотных остатков белков с нуклеотидами ДНК	176

<i>Матвеева О. В., Назипова Н. Н., Огурцов А. Ю., Шабалина С. А.</i> Оптимизированная модель для конструирования эффективных малых РНК, играющих важную роль в регуляции работы генов у животных и растений	178
<i>Панкратов А. Н., Пятков М. И., Руднев В. Р., Куликова Л. И.</i> Быстрое распознавание повторов в генетических последовательностях на основе спектрально-аналитического метода	180
Приложения: экономические системы	182
<i>Боровых Н. И., Красоткина О. В.</i> Применение регуляризованной байесовской логистической регрессии в задаче кредитного скоринга	182
<i>Вальков А. С., Журавлев Ю. И., Каширин Д. О., Рудаков К. В., Стрижов В. В., Корчагин А. Д.</i> Алгоритмы непараметрического прогнозирования объемов грузоперевозок	184
<i>Вальков А. С., Журавлев Ю. И., Каширин Д. О., Рудаков К. В., Стрижов В. В., Корчагин А. Д.</i> Прогнозирование объемов грузоперевозок с учетом экзогенных факторов	186
<i>Дорофеев А. А., Дорофеев Ю. А., Покровская И. В., Чернявский А. Л.</i> Независимая многовариантная экспертиза в задачах исследования слабо формализованных систем управления	188
<i>Дорофеев Ю. А., Покровская И. В., Киселёва Н. Е.</i> Комплекс алгоритмов интеллектуального анализа сложно организованных данных при исследовании слабо формализованных систем управления	190
<i>Шананин А. А.</i> Вычислимая модель железнодорожных грузоперевозок с учетом коммуникационных ограничений	192

Приложения: обработка естественного языка	194
<i>Авдеева Н. В., Лобанова Г. А., Нижулина О. В.</i>	
Классификация фрагментов текста при экспертизе диссертаций на предмет заимствований (плагиата)	194
<i>Варламов М. И., Коршунов А. В.</i>	
Расчет семантической близости концептов на основе кратчайших путей в графе ссылок Википедии	196
<i>Воронцов К. В., Потапенко А. А., Фрей А. И., Апишев М. А., Дойков Н. В., Шапулин А. В., Чиркова Н. А.</i>	
Многокритериальные и многомодальные вероятностные тематические модели коллекций текстовых документов .	198
<i>Гусев В. Д., Мирошниченко Л. А., Саломатина Н. В.</i>	
Структурные аналогии в символьных последовательностях различной языковой природы	200
<i>Дербенёв Н. В., Козлюк Д. А., Никитин В. В., Толчеев В. О.</i>	
Экспериментальное исследование методов выявления нечетких дубликатов научных публикаций	202
<i>Дударенко М. А.</i>	
Регуляризация многоязычных тематических моделей . .	204
<i>Емельянов Г. М., Михайлов Д. В., Козлов А. П.</i>	
Формирование единиц представления знаний в задаче их оценки на основе открытых тестов	206
<i>Жаркова О. С., Берестнева О. Г., Моисеенко А. В., Марухина О. В.</i>	
Компьютерное психологическое тестирование на базе портала MultiTest	208
<i>Мохов А. С., Толчеев В. О.</i>	
Способы повышения точности классификации двуязычных текстовых библиографических документов	210
<i>Чучупал В. Я., Коренчиков А. А.</i>	
Моделирование вариативности произношения для уменьшения уровня ошибок при распознавании речи	212
<i>Юлин С. С., Паламарь И. Н.</i>	
Графическая вероятностная модель со скрытыми состояниями на основе главных многообразий	214

Приложения: геофизика	216
<i>Белозеров Б. В., Бочков А. С., Урмаев О. С., Фукс О. М.</i> Методы машинного обучения при восстановлении обстановки осадконакопления в задачах геостатистического моделирования	216
<i>Воскобойникова Г. М., Кельманов А. В., Хайретдинов М. С.</i> Апостериорные алгоритмы в проблеме геофизического мониторинга	218
<i>Ковалевский В. В., Брагинская Л. П., Григорюк А. П., Кратов С. В.</i> Научная информационная система «Активная сейсмология» для комплексных геофизических исследований . . .	220
<i>Мандрикова О. В., Жижикина Е. А.</i> Оценка состояния геомагнитного поля на основе совмещения вейвлет-преобразования с радиальными нейронными сетями	222
<i>Мандрикова О. В., Заляев Т. Л.</i> Моделирование вариаций космических лучей и выделение аномалий на основе совмещения вейвлет-преобразования с нейронными сетями	224
<i>Чичева М. А., Юзькив Р. Р.</i> Сжатие гиперспектральных данных на основе метода кодирования с преобразованием	226
Приложения: информационная безопасность	228
<i>Браницкий А. А., Тимофеев А. В.</i> Интеллектуальные методы обнаружения и классификации сетевых атак	228
<i>Ручай А. Н., Горшенин В. В., Маткин И. А.</i> Разработка прототипа централизованной системы избирательной многофакторной биометрической аутентификации	230
<i>Сычугов А. А., Нижилин М. С., Середин О. С.</i> Идентификация потенциально опасных клиентских запросов на основе многоклассового распознавания образов	232

Contents

Theory of Classification	8
<i>Abramov V., Seredin O., Mottl V.</i>	
Pattern recognition in arbitrary metric spaces	9
<i>Zhuravlev Yu.</i>	
On a new modification of estimation algorithms	11
<i>Zagoruiko N., Kutnenko O., Zyryanov A., Levanov D.</i>	
Learning to recognition without overfitting	13
<i>Levdik P., Chernousova E., Sungurov D., Mottl V.</i>	
Experimental study of hypothetical cross validation applied to determination of structural parameters in featureless pat- tern recognition	15
<i>Lepskiy A.</i>	
Comparison of distorted histograms by probability methods	17
<i>Makhina G.</i>	
Learning isotonic nearest neighbor classifier from non- isotonic training set	19
<i>Nedelko V.</i>	
Some aspects of estimating a quality of decision functions construction methods	21
<i>Nikishin M., Krasotkina O., Mottl V.</i>	
Multiclass pattern recognition from large training sets	23
<i>Stenina M., Kuznetsov M.</i>	
Partial orders aggregation for preference learning problem .	25
<i>Tatarchuk A., Mottl V.</i>	
Bayesian feature-selective formulations of the pattern recog- nition problem based on support vector machine	27
<i>Turkov P., Krasotkina O.</i>	
Feature selection in the pattern recognition problem under concept drift	29
<i>Chernousova E., Levdik P., Mottl V.</i>	
Nonenumerative cross validation for the determination of structural parameters in feature-selective support vector machines	31

<i>Chulichkov A., Yuan B.</i>	
Estimation of minimum possibility of losses and minimax estimation: A comparative analysis	33
Distance-Based Data Analysis	34
<i>Berikov V.</i>	
Ensemble of clustering algorithms using weights	35
<i>Dvoenko S.</i>	
Bipartial objective function for clustering of a set of elements in terms of pairwise comparisons	37
<i>Dvoenko S., Pshenichny D.</i>	
Optimal correction of metrical violations in matrices of pairwise comparisons	39
<i>Iofina G., Minaev A., Polyakov Yu., Maximov Yu.</i>	
A Study of Metrics for Semisupervised Learning Problems .	41
<i>Pushnyakov A.</i>	
On combinatorial bounds for maximal ε -partitions of a finite metric space	43
<i>Rudakov K.</i>	
On relational factorizations of semimetric cone	45
Rule-Based Classification	46
<i>Genrikhov I.</i>	
About splitting criteria used for synthesis of decision trees .	47
<i>Djukova E., Nikiforov A., Prokofjev P.</i>	
Efficient parallel dualization algorithm	49
<i>Djukova E., Prokofjev P.</i>	
New asymptotically optimal algorithms for dualization . . .	51
<i>Pokrovskaya I., Goldovskaya M., Dorofeyuk Y., Kiseleva N.</i>	
Intellectual processing methods of qualitative data	53
<i>Torshin I., Rudakov K.</i>	
Invariants of labeled graphs in the combinatorial theory of solvability	55

<i>Yankovskaya A., Yamshanov A., Krivdyuk N.</i> Cognitive tools in intelligent systems of decision-making support	57
Time Series Analysis	58
<i>Zagoruiko N., Borisova I., Levanov D.</i> Description of multivariate dynamic processes with concepts hierarchy	59
<i>Zagoruiko N., Tatarnikov V.</i> Use of redundancy in cubes of data for detection of errors and filling of blanks	61
<i>Teklina L., Kotel'nikov I.</i> Synthesis of linear multidimensional systems for the quasi-invariant control using methods of pattern recognition	63
<i>Filipenkov N., Petrova M.</i> On the Analysis of Multidimensional Time Series	65
<i>Chernyavskiy A., Dorofeyuk A., Dorofeyuk Y.</i> Intellectual methods of structural analysis of multidimensional trajectories	67
<i>Yankovskaya A., Ametov R., Kitler S.</i> About creation of self-learning support system for decision-making on stock markets based on fault-tolerant subset of signal features	69
Computational Complexity and Optimization	70
<i>Ageev A., Kel'manov A., Pyatkin A.</i> NP-hardness of the Euclidean MAX-CUT problem	71
<i>Borisov B.</i> Efficient estimation of main frequency instability characteristic of oscillators	73
<i>Volkov Yu., Miroshnichenko V., Salienko A.</i> Mathematical modeling of hill diagram for Kaplan turbine	75
<i>Gasnikov A., Krymova E.</i> Stochastic online optimization in case of heavy-tailed stochastic gradients	77

<i>Gimadi E., Istomin A., Rykov I., Tsidulko O.</i>	
On asymptotical solvability for the m -peripatetic salesman problem on random instances unbounded from above	79
<i>Gimadi E., Kel'manov A., Pyatkin A., Khachay M.</i>	
Approximation algorithms for finding several disjoint cliques in a complete graph with minimal total weight	81
<i>Glinsky B., Marchenko M., Rodionov A., Karavaev D., Podkorytov D.</i>	
Mappings of parallel algorithms on supercomputers with ekzaflops performance on the basis of simulation	83
<i>Gornov A., Zarodnyuk T.</i>	
Computing technology for estimation of convexity degree of the multiextremal function	85
<i>Iofina G., Maximov Yu.</i>	
Computational Complexity of Unbalanced Boolean Minimization within Depth-2 Circuits	87
<i>Karatsuba E.</i>	
Fast algorithms. Project presentation	89
<i>Kel'manov A.</i>	
Efficient approximation algorithms with performance guarantees for some discrete optimization problems in analysis and recognition of sequences	91
<i>Kel'manov A., Khamidullin S.</i>	
Efficient approximation algorithm for a sequence partitioning problem	93
<i>Kel'manov A., Khandeev V.</i>	
An exact pseudopolynomial algorithm for a vectors set bipartitioning problem	95
<i>Kirsh D., Kupriyanov A.</i>	
Crystal lattice identification based on the estimation of Bravais and Wigner–Seitz unit cell parameters	97
<i>Khachay M., Neznakhina E.</i>	
Polynomial-time approximability of geometric multiple traveling salesmen problem	99
<i>Chernov V.</i>	
Ternary computer arithmetic: History, problems, applications	101

<i>Chykhradze K., Korshunov A., Kuzyurin N.</i>	
On a model of social network with user communities for distributed generation of random social graphs	103
Image Processing	104
<i>Gracheva I., Kopylov A., Krasotkina O.</i>	
Adaptive procedure for generalized smoothing of images on the basis of statistical approach	105
<i>Karkishchenko A., Mnukhin V.</i>	
Symmetrization of the image points defined by statistical sampling	107
<i>Larin A.</i>	
The development of three-dimensional human body scanner for virtual fitting room using Kinect	109
<i>Fedotov N., Syemov A., Moiseev A.</i>	
Intelligent capabilities hypertrace transform: constructing features with predetermined properties	111
Image Analysis	112
<i>Egorov A., Vishnyakov B.</i>	
Method of semiautomatic contextual scene markup	113
<i>Murashov D., Berezin A., Ivanova E.</i>	
Comparing images of paintings using informative fragments	115
<i>Kiy K.</i>	
A new real-time method of contextual image description and its application in robot navigation and intelligent control . .	117
<i>Kornilov F.</i>	
Structural differences detection on the images	119
<i>Karkishchenko A., Mnukhin V.</i>	
Topological filtration for digital images recognition and symmetry analysis	121
Video Analysis	122
<i>Karateev S., Kostromov N., Beketova I., Vizilter Yu.</i>	
The complex information processing based on U-gramms for error correction in recognized video sequences of symbols . .	123

<i>Karpov R., Kopylov A.</i> Clustering for digital video stabilization	125
<i>Lebedev M., Bondarenko M., Komarov D., Stepaniants D., Vy- golov O., Vizilter Yu., Zheltov S.</i> An algorithm for real and synthetic images matching with application to aviation combined vision system	127
<i>Mitekin V., Fedoseev V.</i> A new method for high-capacity information hiding in video robust against temporal desynchronization	129
<i>Morozov A., Vaish A., Polupanov A., Antsiperov V., Ly- chkov I., Alfimtsev A., Deviatkov V.</i> Development of concurrent object-oriented logic program- ming platform for the intelligent monitoring of anomalous human activities	131
<i>Panov A.</i> Algebraic properties for recognition operators in modeling visual perception of dynamic scenes	133
<i>Tashlinskiy A., Voronov S.</i> On the use of mutual information as objective function for image parameters' estimation	135
<i>Chernousov V., Savchenko A.</i> Dependence of the moving forklift truck detection morpho- logical algorithm accuracy on the intensity of the additive noise	137
Image Recognition	138
<i>Beketova I., Karateev S., Ososkov M., Kostromov N.</i> Technology standard picture person using hardware-software complex preparation and control of digital images and 3D- model of the surface faces	139
<i>Kushnir O., Seredin O.</i> Applying the procedure of binary images comparison based on skeletons to the task of medicinal leaves recognition . . .	141
<i>Lange M., Ganebnykh S.</i> An efficiency of hierarchical algorithms for two-dimensional object classification in terms of fidelity-complexity ratio . .	143

<i>Matveev I., Novik V.</i>	
Selecting best iris template from sequence	145
Images Shape Analysis	146
<i>Vizilter Yu., Zheltov S.</i>	
Similarity measures and comparison metrics for image shapes	147
<i>Gorbatsevich V., Vizilter Yu., Zheltov S., Rubis A., Vorotnikov A.</i>	
MORPHLETS: A New Class of Tree-Structured Morphological Descriptors of Image Shape Based on Haar-like Wavelets	149
<i>Mestetskiy L.</i>	
Medial width of figures — an image shape descriptor	151
<i>Pyt'ev Yu.</i>	
Oblique projectors, relative forms and subjective models in image morphology	153
<i>Sidyakin S., Vizilter Yu.</i>	
Morphological shape descriptors of binary images based on elliptical structuring elements	155
<i>Stepanov A., Seredin O., Kushnir O.</i>	
Automatic selection of the most appropriate values of pruning, approximation, and merging parameters for binary images skeletons	157
Applications: Medicine	158
<i>Antsiperov V.</i>	
Cortical rhythms detection based on the values of electroencephalography analytic spectra correlation in specific frequency ranges	159
<i>Berestneva O., Volovodenko V., Gerget O., Sharopin K., Osadhaya I.</i>	
Multidimensional medical data visualization methods based on generalized graphic images	161
<i>Makarov V., Beklaryan L., Belousov F.</i>	
Steady-state regimens in Henning model and its modifications	163
<i>Oplachko E., Rykunov S., Ustinin M.</i>	
Cloud service MathBrain to analyze encephalography data .	165

<i>Pankratova N., Ustinin M.</i> Spectral analysis of the high frequency region of magnetic encephalography data	167
<i>Rykunov S., Sychev V., Ustinin M.</i> Extraction and localization of the sources of the brain auditory cortical response from the noisy data of magneto- and electroencephalography	169
<i>Sushkova O., Korolev M., Gabova A., Illarioshkin S., Karabanov A., Obukhov K., Obukhov Y.</i> Joint analysis of the signals of electroencephalograms, electromyograms, and mechanical tremor of the Parkinson's disease in the early stage	171
<i>Uspenskiy V., Vorontsov K., Tselykh V., Bunakov V., Efimova I., Polezhaev V.</i> Informational analysis of electro cardio signals for multidisease diagnostic system	173
<i>Ustinin M.</i> Functional tomography of the human brain based on frequency-pattern analysis of the encephalography data	175
Applications: Bioinformatics	176
<i>Kuznetsov E., Anashkina A., Esipova N., Tumanyan V.</i> Cluster analysis of spatial contacts of amino acid residues of proteins with DNA nucleotides	177
<i>Matveeva O., Nazipova N., Ogurtsov A., Shabalina S.</i> Optimized model for the design of effective small RNAs that play an important role in regulation of gene expression in animals and plants	179
<i>Pankratov A., Pyatkov M., Rudnev V., Kulikova L.</i> Rapid recognition of repeats in genetic sequences based on the spectral-analytical method	181
Applications: Economics	182
<i>Borovykh N., Krasotkina O.</i> Application of the regularized bayessian logistic regression to the credit scoring problem	183

<i>Valkov A., Zhuravlev Yu., Kashirin D., Rudakov K., Strijov V., Korchagin A.</i> Nonparametric algorithms for forecasting cargo transportation quantities	185
<i>Valkov A., Zhuravlev Yu., Kashirin D., Rudakov K., Strijov V., Korchagin A.</i> Forecasting cargo transportation quantities with regard to exogenous factors	187
<i>Dorofeyuk A., Dorofeyuk Y., Pokrovskaya I., Chernyavskiy A.</i> Independent multivariate expertise in the weakly formalized management systems research	189
<i>Dorofeyuk Y., Pokrovskaya I., Kiseleva N.</i> The complicated data mining algorithms complex in the study of weakly formalized management systems	191
<i>Shananin A.</i> Rail freight computational model under communication constraints.	193
Applications: Natural Language Processing	194
<i>Avdeeva N., Lobanova G., Nikulina O.</i> Classification of text fragments in the examination of theses on the subject of borrowing (plagiarism)	195
<i>Varlamov M., Korshunov A.</i> Computing semantic similarity of concepts using shortest paths in Wikipedia link graph	197
<i>Vorontsov K., Potapenko A., Frei O., Apishev M., Doikov N., Shapulin A., Chirkova N.</i> Multi-criteria and multi-modal probabilistic topic models of text collections	199
<i>Gusev V., Miroshnichenko L., Salomatina N.</i> Structural analogies in symbolic sequences of different nature	201
<i>Derbenev N., Kozliuk D., Nikitin V., Tolcheev V.</i> Experimental research of near-duplicate detection methods for scientific papers	203
<i>Dudarenko M.</i> Multilanguage topic models regularization	205

<i>Emelyanov G., Mikhaylov D., Kozlov A.</i>	
Forming of the knowledge units' representations to estimate them by open-form tests	207
<i>Zharkova O., Berestneva O., Moiseenko A., Marukhina O.</i>	
Psychological computer testing based on MultiTest portal	209
<i>Mokhov A., Tolcheev V.</i>	
Methods of bilingual text documents classification precision increasing	211
<i>Chuchupal V., Korenchikov A.</i>	
Improving speech recognition accuracy by means of word pronunciation modeling	213
<i>Yulin S., Palamar I.</i>	
Probabilistic graphical model with hidden states based on the principal manifolds	215
Applications: Geophysics	216
<i>Belozеров B., Bochkov A., Ushmaev O., Fuks O.</i>	
Machine learning for sedimentation environment study in the tasks of geostatistics simulation	217
<i>Voskoboynikova G., Kel'manov A., Khairtdinov M.</i>	
Posteriori algorithms in a problem of geophysical monitoring	219
<i>Kovalevsky V., Braginskaya L., Grigoruk A., Kratov S.</i>	
Scientific Information System "Active Seismology" for integrated geophysical studies	221
<i>Mandrikova O., Zhizhikina E.</i>	
Estimation of degree of the geomagnetic field disturbance based on the combined use of wavelet transform with radial neural networks	223
<i>Mandrikova O., Zalyaev T.</i>	
Modeling of cosmic ray variations and allocation of anomalies based on a combination of wavelet transform with neural networks	225
<i>Chicheva M., Yuzkiv R.</i>	
Hyperspectral data compression based on transform coding method	227

Applications: Information Security	228
<i>Branitskiy A., Timofeev A.</i>	
Intellectual methods of detection and classification of network attacks	229
<i>Ruchay A., Gorshenin V., Matkin I.</i>	
Development of prototype centralized system of election multifactor biometric authentication	231
<i>Sychugov A., Nikishin M., Seredin O.</i>	
Detection of potentially dangerous client inquiries on the basis of multiclass pattern recognition	233

Авторский указатель**А**

Абрамов В. И. 8
Авдеева Н. В. 194
Агеев А. А. 70
Алфимцев А. Н. 130
Аметов Р. В. 68
Анашкина А. А. 176
Анциперов В. Е. ... 130, 158
Апишев М. А. 198

Б

Бекетова И. В. 122, 138
Бекларян Л. А. 162
Белозеров Б. В. 216
Белоусов Ф. А. 162
Березин А. В. 114
Берестнева О. Г. ... 160, 208
Бериков В. Б. 34
Бондаренко М. А. 126
Борисов Б. Д. 72
Борисова И. А. 58
Борových Н. И. 182
Бочков А. С. 216
Брагинская Л. П. 220
Браницкий А. А. 228
Бунаков В. А. 172

В

Вайш А. 130
Вальков А. С. 184, 186
Варламов М. И. 196
Визильтер Ю. В. .. 122, 126,
146, 148, 154
Вишняков Б. В. 112
Волков Ю. С. 74

Воловоденко В. А. 160
Воронов С. В. 134
Воронцов К. В. 172, 198
Воротников А. В. 148
Воскобойникова Г. М. ... 218
Выголов О. В. 126

Г

Габова А. В. 170
Ганебных С. Н. 142
Гасников А. В. 76
Генрихов И. Е. 46
Гергет О. М. 160
Гимади Э. Х. 78, 80
Глинский Б. М. 82
Гольдовская М. Д. 52
Горбацевич В. С. 148
Горнов А. Ю. 84
Горшенин В. В. 230
Грачева И. А. 104
Григорюк А. П. 220
Гусев В. Д. 200

Д

Двоенко С. Д. 36, 38
Девятков В. В. 130
Дербенёв Н. В. 202
Дойков Н. В. 198
Дорофеев А. А. 66, 188
Дорофеев Ю. А. 52, 66, 188,
190
Дударенко М. А. 204
Дюкова Е. В. 48, 50

Е

Егоров А. И. 112

Емельянов Г. М. 206
 Есипова Н. Г. 176
 Ефимова И. В. 172

Ж

Жаркова О. С. 208
 Желтов С. Ю. 126, 146, 148
 Жижикина Е. А. 222
 Журавлев Ю. И. ... 10, 184,
 186

З

Загоруйко Н. Г. ... 12, 58, 60
 Заляев Т. Л. 224
 Зароднюк Т. С. 84
 Зырянов А. О. 12

И

Иванова Е. Ю. 114
 Иллариошкин С. Н. 170
 Иофина Г. В. 40, 86
 Истомин А. М. 78

К

Карабанов А. В. 170
 Караваев Д. А. 82
 Каратеев С. Л. 122, 138
 Карацуба Е. А. 88
 Каркищенко А. Н. . 106, 120
 Карпов Р. О. 124
 Каширин Д. О. 184, 186
 Кельманов А. В. 218
 Кельманов А. В. . 70, 80, 90,
 92, 94
 Кий К. И. 116
 Кириш Д. В. 96
 Киселёва Н. Е. 52, 190
 Китлер С. В. 68
 Ковалевский В. В. 220

Козлов А. П. 206
 Козлюк Д. А. 202
 Комаров Д. В. 126
 Кошылов А. В. 104, 124
 Коренчиков А. А. 212
 Корнилов Ф. А. 118
 Королев М. С. 170
 Корчагин А. Д. 184, 186
 Коршунов А. В. 102, 196
 Костромов Н. А. ... 122, 138
 Котельников И. В. 62
 Красоткина О. В. 22, 28,
 104, 182
 Кратов С. В. 220
 Кривдюк Н. М. 56
 Крымова Е. А. 76
 Кузнецов Е. Н. 176
 Кузнецов М. П. 24
 Кузюрин Н. Н. 102
 Куликова Л. И. 180
 Куприянов А. В. 96
 Кутненко О. А. 12
 Кушнир О. А. 140, 156

Л

Ланге М. М. 142
 Ларин А. О. 108
 Лебедев М. А. 126
 Леванов Д. А. 12, 58
 Левдик П. В. 14, 30
 Лепский А. Е. 16
 Лобанова Г. А. 194
 Лычков И. И. 130

М

Макаров В. Л. 162
 Максимов Ю. В. 40, 86
 Мандрикова О. В. .. 222, 224

Марухина О. В. 208
Марченко М. А. 82
Матвеев И. А. 144
Матвеева О. В. 178
Маткин И. А. 230
Махина Г. А. 18
Местецкий Л. М. 150
Минаев А. В. 40
Мирошниченко В. Л. 74
Мирошниченко Л. А. 200
Митекин В. А. 128
Михайлов Д. В. 206
Мнухин В. Б. 106, 120
Моисеев А. В. 110
Моисеенко А. В. 208
Морозов А. А. 130
Моттль В. В. 8, 14, 22, 26, 30
Мохов А. С. 210
Мурашов Д. М. 114

Н

Назипова Н. Н. 178
Неделько В. М. 20
Незнахина Е. Д. 98
Никитин В. В. 202
Никифоров А. Г. 48
Никишин М. С. 22, 232
Никулина О. В. 194
Новик В. П. 144

О

Обухов К. Ю. 170
Обухов Ю. В. 170
Огурцов А. Ю. 178
Оплачко Е. С. 164
Осадчая И. А. 160
Ососков М. В. 138

П

Паламарь И. Н. 214
Панкратов А. Н. 180
Панкратова Н. М. 166
Панов А. И. 132
Петрова М. А. 64
Подкорытов Д. И. 82
Покровская И. В. ... 52, 188,
190
Полежаев В. А. 172
Полупанов А. Ф. 130
Поляков Ю. С. 40
Потапенко А. А. 198
Прокофьев П. А. 48, 50
Пушняков А. С. 42
Пшеничный Д. О. 38
Пытьев Ю. П. 152
Пяткин А. В. 70, 80
Пятков М. И. 180

Р

Родионов А. С. 82
Рубис А. Ю. 148
Рудаков К. В. ... 44, 54, 184,
186
Руднев В. Р. 180
Ручай А. Н. 230
Рыков И. А. 78
Рыкунов С. Д. 164, 168

С

Сёмов А. А. 110
Савченко А. В. 136
Салиенко А. Е. 74
Саломатина Н. В. 200
Середин О. С. ... 8, 140, 156,
232
Сидякин С. В. 154

Стенина М. М. 24
 Степанов А. В. 156
 Степаньянц Д. Г. 126
 Стрижов В. В. 184, 186
 Сунгуров Д. С. 14
 Сушкова О. С. 170
 Сычев В. В. 168
 Сычугов А. А. 232

Т

Татарников В. В. 60
 Татарчук А. И. 26
 Ташлинский А. Г. 134
 Теклина Л. Г. 62
 Тимофеев А. В. 228
 Толчеев В. О. 202, 210
 Торшин И. Ю. 54
 Туманян В. Г. 176
 Турков П. А. 28

У

Успенский В. М. 172
 Устинин М. Н. 164, 166, 168,
 174
 Ушмаев О. С. 216

Ф

Федосеев В. А. 128
 Федотов Н. Г. 110
 Филипенков Н. В. 64
 Фрей А. И. 198
 Фукс О. М. 216

Х

Хайретдинов М. С. 218

Хамидуллин С. А. 92
 Хандеев В. И. 94
 Хачай М. Ю. 80, 98

Ц

Целых В. Р. 172
 Цидулко О. Ю. 78

Ч

Чернов В. М. 100
 Черноусов В. О. 136
 Черноусова Е. О. 14, 30
 Чернявский А. Л. ... 66, 188
 Чиркова Н. А. 198
 Чихрадзе К. К. 102
 Чичева М. А. 226
 Чуличков А. И. 32
 Чучупал В. Я. 212

Ш

Шабалина С. А. 178
 Шананин А. А. 192
 Шапулин А. В. 198
 Шаропин К. А. 160

Ю

Юань Б. 32
 Юзькив Р. Р. 226
 Юлин С. С. 214

Я

Ямпанов А. В. 56
 Янковская А. Е. 56, 68

Author index

A

Abramov V. 9
Ageev A. 71
Alfimtsev A. 131
Ametov R. 69
Anashkina A. 177
Antsiperov V. 131, 159
Apishev M. 199
Avdeeva N. 195

B

Beketova I. 123, 139
Beklaryan L. 163
Belousov F. 163
Belozеров B. 217
Berestneva O. 161, 209
Berezin A. 115
Berikov V. 35
Bochkov A. 217
Bondarenko M. 127
Borisova I. 59
Borisov B. 73
Borovykh N. 183
Braginskaya L. 221
Branitskiy A. 229
Bunakov V. 173

C

Chernousova E. 15, 31
Chernousov V. 137
Chernov V. 101
Chernyavskiy A. 67, 189
Chicheva M. 227
Chirkova N. 199
Chuchupal V. 213
Chulichkov A. 33

Chykhhradze K. 103

D

Derbenev N. 203
Deviatkov V. 131
Djukova E. 49, 51
Doikov N. 199
Dorofeyuk A. 67, 189
Dorofeyuk Y. 53, 67, 189, 191
Dudarenko M. 205
Dvoenko S. 37, 39

E

Efimova I. 173
Egorov A. 113
Emelyanov G. 207
Esipova N. 177

F

Fedoseev V. 129
Fedotov N. 111
Filipenkov N. 65
Frei O. 199
Fuks O. 217

G

Gabova A. 171
Ganebnykh S. 143
Gasnikov A. 77
Genrikhov I. 47
Gerget O. 161
Gimadi E. 79, 81
Glinsky B. 83
Goldovskaya M. 53
Gorbatsevich V. 149
Gornov A. 85

Gorshenin V. 231
 Gracheva I. 105
 Grigoruk A. 221
 Gusev V. 201

I

Illarioshkin S. 171
 Iofina G. 41, 87
 Istomin A. 79
 Ivanova E. 115

K

Karabanov A. 171
 Karateev S. 123, 139
 Karatsuba E. 89
 Karavaev D. 83
 Karkishchenko A. ... 107, 121
 Karpov R. 125
 Kashirin D. 185, 187
 Kel'manov A. . 71, 81, 91, 93,
 95, 219
 Khachay M. 81, 99
 Khairtdinov M. 219
 Khamidullin S. 93
 Khandeev V. 95
 Kirsh D. 97
 Kiseleva N. 53, 191
 Kitler S. 69
 Kiy K. 117
 Komarov D. 127
 Kopylov A. 105, 125
 Korchagin A. 185, 187
 Korenchikov A. 213
 Kornilov F. 119
 Korolev M. 171
 Korshunov A. 103, 197
 Kostromov N. 123, 139
 Kotel'nikov I. 63

Kovalevsky V. 221
 Kozliuk D. 203
 Kozlov A. 207
 Krasotkina O. ... 23, 29, 105,
 183
 Kratov S. 221
 Krivdyuk N. 57
 Krymova E. 77
 Kulikova L. 181
 Kupriyanov A. 97
 Kushnir O. 141, 157
 Kutnenko O. 13
 Kuznetsov E. 177
 Kuznetsov M. 25
 Kuzyurin N. 103

L

Lange M. 143
 Larin A. 109
 Lebedev M. 127
 Lepskiy A. 17
 Levanov D. 13, 59
 Levdik P. 15, 31
 Lobanova G. 195
 Lychkov I. 131

M

Makarov V. 163
 Makhina G. 19
 Mandrikova O. 223, 225
 Marchenko M. 83
 Marukhina O. 209
 Matkin I. 231
 Matveeva O. 179
 Matveev I. 145
 Maximov Yu. 41, 87
 Mestetskiy L. 151
 Mikhaylov D. 207

Minaev A.41
Miroshnichenko L.201
Miroshnichenko V.75
Mitekin V.129
Mnukhin V.107, 121
Moiseenko A.209
Moiseev A.111
Mokhov A.211
Morozov A.131
Mottl V.9, 15, 23, 27, 31
Murashov D.115

N

Nazipova N.179
Nedelko V.21
Neznakhina E.99
Nikiforov A.49
Nikishin M.23, 233
Nikitin V.203
Nikulina O.195
Novik V.145

O

Obukhov K.171
Obukhov Y.171
Ogurtsov A.179
Oplachko E.165
Osadhaya I.161
Ososkov M.139

P

Palamar I.215
Pankratova N.167
Pankratov A.181
Panov A.133
Petrova M.65
Podkorytov D.83
Pokrovskaya I.53, 189, 191
Polezhaev V.173

Polupanov A.131
Polyakov Yu.41
Potapenko A.199
Prokofjev P.49, 51
Pshenichny D.39
Pushnyakov A.43
Pyatkin A.71, 81
Pyatkov M.181
Pyt'ev Yu.153

R

Rodionov A.83
Rubis A.149
Ruchay A.231
Rudakov K.45, 55, 185, 187
Rudnev V.181
Rykov I.79
Rykunov S.165, 169

S

Salienko A.75
Salomatina N.201
Savchenko A.137
Seredin O.9, 141, 157, 233
Shabalina S.179
Shananin A.193
Shapulin A.199
Sharopin K.161
Sidyakin S.155
Stenina M.25
Stepanians D.127
Stepanov A.157
Strijov V.185, 187
Sungurov D.15
Sushkova O.171
Sychev V.169
Sychugov A.233
Syemov A.111

T

Tashlinskiy A.135
 Tatarchuk A.27
 Tatarnikov V.61
 Teklina L.63
 Timofeev A.229
 Tolcheev V.203, 211
 Torshin I.55
 Tselykh V.173
 Tsidulko O.79
 Tumanyan V.177
 Turkov P.29

U

Ushmaev O.217
 Uspenskiy V.173
 Ustinin M. 165, 167, 169, 175

V

Vaish A.131
 Valkov A.185, 187
 Varlamov M.197
 Vishnyakov B.113
 Vizilter Yu. ...123, 127, 147,
 149, 155

Volkov Yu.75
 Volovodenko V.161
 Voronov S.135
 Vorontsov K.173, 199
 Vorotnikov A.149
 Voskoboynikova G.219
 Vygolov O.127

Y

Yamshanov A.57
 Yankovskaya A.57, 69
 Yuan B.33
 Yulin S.215
 Yuzkiv R.227

Z

Zagoruiko N.13, 59, 61
 Zalyaev T.225
 Zarodnyuk T.85
 Zharkova O.209
 Zheltov S.127, 147, 149
 Zhizhikina E.223
 Zhuravlev Yu. ...11, 185, 187
 Zyryanov A.13

Научное издание

ИНТЕЛЛЕКТУАЛИЗАЦИЯ ОБРАБОТКИ ИНФОРМАЦИИ

Тезисы докладов
10-й международной конференции ИОИ-2014

Напечатано с готового оригинал-макета

Сдано в набор 11.08.14. Подписано в печать 18.09.14.
Формат 60×90/16. Бумага офсетная. Печать цифровая.
Усл.-печ. л. 16,63. Уч.-изд. л. 12,5. Тираж 200 экз.

Заказ № 6021.

Издательство «ТОРУС ПРЕСС»
121614, г. Москва, ул. Крылатская 29-1-43
e-mail: torus@torus-press.ru
<http://www.torus-press.ru>

Отпечатано в цифровой типографии «Буки Веди»
на оборудовании Konica Minolta
105066, г. Москва, ул. Новорязанская, д. 38, стр. 1, пом. IV
Тел.: (495) 926-63-96, www.bukivedi.com, info@bukivedi.com

