

Communications in Computer and Information Science

661

Commenced Publication in 2007

Founding and Former Series Editors:

Alfredo Cuzzocrea, Dominik Ślęzak, and Xiaokang Yang

Editorial Board

Simone Diniz Junqueira Barbosa

*Pontifical Catholic University of Rio de Janeiro (PUC-Rio),
Rio de Janeiro, Brazil*

Phoebe Chen

La Trobe University, Melbourne, Australia

Xiaoyong Du

Renmin University of China, Beijing, China

Joaquim Filipe

Polytechnic Institute of Setúbal, Setúbal, Portugal

Orhun Kara

TÜBİTAK BİLGEM and Middle East Technical University, Ankara, Turkey

Igor Kotenko

*St. Petersburg Institute for Informatics and Automation of the Russian
Academy of Sciences, St. Petersburg, Russia*

Ting Liu

Harbin Institute of Technology (HIT), Harbin, China

Krishna M. Sivalingam

Indian Institute of Technology Madras, Chennai, India

Takashi Washio

Osaka University, Osaka, Japan


More information about this series at <http://www.springer.com/series/7899>

Dmitry I. Ignatov · Mikhail Yu. Khachay
Valeri G. Labunets · Natalia Loukachevitch
Sergey I. Nikolenko · Alexander Panchenko
Andrey V. Savchenko · Konstantin Vorontsov (Eds.)

Analysis of Images, Social Networks and Texts

5th International Conference, AIST 2016
Yekaterinburg, Russia, April 7–9, 2016
Revised Selected Papers

Editors

Dmitry I. Ignatov 
National Research University Higher School
of Economics
Moscow
Russia

Mikhail Yu. Khachay
Krasovsky Institute of Mathematics and
Mechanics
Yekaterinburg
Russia

Valeri G. Labunets
Ural Federal University
Yekaterinbug
Russia

Natalia Loukachevitch
Research Computing Center
Lomonosov Moscow State University
Moscow
Russia

Sergey I. Nikolenko
National Research University Higher School
of Economics
St. Petersburg
Russia

Alexander Panchenko
Technische Universität Darmstadt
Darmstadt
Germany

Andrey V. Savchenko
Laboratory of Algorithms and Technologies
for Networks Analysis
National Research University Higher School
of Economics
Nizhny Novgorod
Russia

Konstantin Vorontsov
Dorodnicyn Computing Centre of Russian
Academy of Sciences
Moscow
Russia

ISSN 1865-0929

Communications in Computer and Information Science

ISBN 978-3-319-52919-6

DOI 10.1007/978-3-319-52920-2

ISSN 1865-0937 (electronic)

ISBN 978-3-319-52920-2 (eBook)

Library of Congress Control Number: 2017930145

© Springer International Publishing AG 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature

The registered company is Springer International Publishing AG

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume contains proceedings of the 5th International conference on Analysis of Images, Social Networks, and Texts (AIST 2016)¹. The previous conferences during 2012–2015 attracted a significant number of students, researchers, academics, and engineers working on interdisciplinary data analysis of images, texts, and social networks.

The broad scope of AIST makes it an event where researchers from different domains, such as image and text processing, exploiting various data analysis techniques, can meet and exchange ideas. We strongly believe that this may lead to crossfertilisation of ideas between researchers relying on modern data analysis machinery. Therefore, AIST brings together all kinds of applications of data-mining and machine learning techniques. The conference allows specialists from different fields to meet each other, present their work, and discuss both theoretical and practical aspects of their data analysis problems. Another important aim of the conference is to stimulate scientists and people from industry to benefit from the knowledge exchange and identify possible grounds for fruitful collaboration.

The conference was held during April 7–9, 2016. Following an already established tradition, the conference was organised in Yekaterinburg, a cross-roads between European and Asian parts of Russia, the capital of the Urals region. The key topics of AIST are analysis of images and videos; natural language processing and computational linguistics; social network analysis; pattern recognition, machine learning and data mining; recommender systems and collaborative technologies; Semantic Web, ontologies and their applications.

The Program Committee and the reviewers of the conference included well-known experts in data mining and machine learning, natural language processing, image processing, social network analysis, and related areas from leading institutions of 27 countries including Australia, Bangladesh, Belgium, Brazil, Croatia, Cyprus, Egypt, Estonia, Finland, France, Germany, Greece, India, Ireland, Israel, Italy, Japan, Lithuania, Norway, Portugal, Qatar, Russia, Spain, Switzerland, The Netherlands, UK, Ukraine, and USA.

This year we have received 142 submissions mostly from Russia but also from Australia, Austria, France, Hungary, India, Italy, Mexico, Norway, Sweden, UK, USA, and Vietnam.

Out of 142 submissions only 30 papers were accepted as regular oral papers. Thus, the acceptance rate of this volume was around 21%. In order to encourage young practitioners and researchers, we included three papers by industry speakers to the main volume and 38 papers to the supplementary proceedings. Each submission was reviewed by at least three reviewers, experts in their fields, in order to supply detailed and helpful comments.

¹ <http://aistconf.org/>.

The conference featured several invited talks and an industry session dedicated to current trends and challenges.

The invited talks were:

- Mehdi Kaytoue (Université de Lyon, CNRS, INSA-Lyon, France), “Finding Duplicate Labels in Behavioral Data: An Application for E-Sport Analytics”
- Attila Kertesz-Farkas (Higher School of Economics, Russia), “False Discovery Rate Control for Database Search Methods over Heterogeneous Biological Data”
- Radhakrishnan Delhibabu (Kazan Federal University, Russia), “Analysis of Anisotropic Diffusion in Image Processing”

The business speakers also covered a wide variety of topics². We have included three of these invited talks in the main volume:

- Denis Nikiforov, Alexander Korchagin, and Ruslan Sivakov (Centre of Information Technology, Ekaterinburg, Russia) “An Ontology-Driven Approach to Electronic Document Structure Design”
- Tatyana Prisyach (Speech Technology Center, St. Petersburg, Russia), Valentin Mendeleev, and Dmitry Ubskiy (ITMO-University, St. Petersburg, Russia), “Data Augmentation for Training of Noise Robust Acoustic Models”
- Alexander Semenov (Higher School of Economics, Russia) and Peter Romov (Yandex, Russia), “Performance of Machine Learning Algorithms in Predicting Game Outcome from Drafts in Dota”

We would also like to mention the best conference papers selected by the Program Committee within each section:

- “On Complexity of Searching a Subset of Vectors with Shortest Average under a Cardinality Restriction” by Anton Ereemeev, Alexander Kelmanov, and Artem Pyatkin (*Machine Learning and Data Mining*)
- “The Problem of The Optimal Packing of The Equal Radius Circles for Non-Euclidean Metric” by Alexander Kazakov, Anna Lempert and Huy Liem Nguyen (*Data Analysis, Social Networks and Complex Data*)
- “Parallel Non-blocking Deterministic Algorithm for Online Topic Modeling” by Oleksandr Frei and Murat Apishev (*Natural Language Processing*)
- “Image Processing Algorithms with Structure Transferring Properties on the Basis of Gamma-Normal Model” by Inessa Gracheva and Andrey Kopylov (*Analysis of Images and Video*)

The final round of the SNA Hackathon 2016³ co-organised by OK.Ru and the AIST team took place in conjunction with the main conference during April 6–8.

We would like to thank the authors for submitting their papers and the members of the Program Committee for their efforts in providing exhaustive reviews. We would

² The detailed program of AIST 2016 Business Day can be found at a separate website: <http://dataconf.org>.

³ <http://snahackathon.org/>.

also like to express special gratitude to all the invited speakers and industry representatives.

We deeply thank all the partners and sponsors. Our golden sponsors are Exactpro⁴ and OK.Ru⁵. Exactpro, a fully owned subsidiary of the London Stock Exchange Group, specialises in quality assurance for exchanges, investment banks, brokers, and other financial sector organisations. OK.Ru, part of the Mail.Ru Group, is one of the two largest Russian social networking services. Our bronze sponsor is the Centre of Information Technologies (Centre IT)⁶. This is a Russian engineering company that specializes in creating exclusive IT solutions and services for the interdepartmental electronic interaction both on national and international levels.

We would like to acknowledge the Scientific Fund of Higher School of Economics for providing AIST participants with travel grants. Our special thanks goes to Springer for their help, starting from the first conference call to the final version of the proceedings. Last but not least, we are grateful to all organisers, especially to Eugeniya Vlasova, and the volunteers, whose endless energy saved us at the most critical stages of the conference preparation.

Traditionally, we would like to mention the Russian word “aist” is more than just a simple abbreviation (in Cyrillic) – it means a “stork.” Since it is a wonderful free bird, a symbol of happiness and peace, this stork brought us the inspiration to organise the AIST conference. So we believe that this young and rapidly growing conference will likewise be bringing inspiration to data scientists around the world!

October 2016

Dmitry Ignatov
Michael Khachay
Valery Labunets
Natalia Loukachevitch
Sergey Nikolenko
Alexander Panchenko
Andrey Savchenko
Konstantin Vorontsov

⁴ exactprosystems.com.

⁵ OK.ru.

⁶ centre-it.com.

Organisation

Program Committee Chairs

Dmitry I. Ignatov	National Research University Higher School of Economics, Moscow, Russia
Mikhail Khachay	Krasovsky Institute of Mathematics and Mechanics of Ural Branch of Russian Academy of Sciences, Russia
Valery G. Labunets Natalia Loukachevitch	Ural Federal University, Russia Computing Centre of Lomonosov Moscow State University, Russia
Sergey I. Nikolenko	National Research University Higher School of Economics and Steklov Institute of Mathematics of Russian Academy of Sciences, St. Petersburg, Russia
Alexander Panchenko	Technische Universität Darmstadt, Germany and Université catholique de Louvain, Belgium
Andrey Savchenko	National Research University Higher School of Economics, Nizhny Novgorod, Russia
Konstantin Vorontsov	Dorodnicyn Computing Centre of Russian Academy of Sciences, Russia

Proceedings Chair

Dmitry I. Ignatov	National Research University Higher School of Economics, Russia
-------------------	---

Organising Committee

Anna Golubtsova	National Research University Higher School of Economics, Moscow
Irina Dolgaleva	National Research University Higher School of Economics, Moscow
Eugeniya Vlasova	National Research University Higher School of Economics, Moscow
Natalia Papulovskaya Ekaterina Borovitina	Ural Federal University, Yekaterinburg Chelyabinsk State University, Russia

Volunteers

Maxim Pasyнков	Krasovsky Institute of Mathematics and Mechanics of Ural Branch of Russian Academy of Sciences, Russia, Yekaterinburg
Anna Voronova	Yandex, Moscow
Eugene Tsymbalov	Webgames and National Research University Higher School of Economics, Moscow
Daria Baranetskaya	National Research University Higher School of Economics, Moscow
Valeria Bubnova	National Research University Higher School of Economics, Moscow

Business Day Chair

Evgenia Vlasova	National Research University Higher School of Economics, Moscow
-----------------	--

Program Committee

Mikhail Ageev	Lomonosov Moscow State University, Russia
Nickolay Arefyev	Lomonosov Moscow State University and Digital Society Lab, Russia
Jaume Baixeries	Universitat Politècnica de Catalunya, Spain
Artem Baklanov	Krasovskii Institute of Mathematics and Mechanics of Ural Branch of the Russian Academy of Sciences, Russia and International Institute for Applied Systems Analysis, Austria
Pedro Balage	University of São Paulo, Brazil
Sergey Bartunov	National Research University Higher School of Economics, Russia and Deep Mind, UK
Malay Bhattacharyya	Indian Institute of Engineering Science and Technology, India
Elena Bolshakova	Moscow State Lomonosov University, Russia
Anastasia Bonch- Osmolovskaya	National Research University Higher School of Economics, Russia
Aurélien Bossard	Université Paris 8, France
Jean-Leon Bouraoui	Université catholique de Louvain, Belgium
Leonid Boytsov	Carnegie Mellon University, USA
Pavel Braslavski	Ural Federal University/Kontur Labs, Russia
Evgeny Burnaev	Institute for Information Transmission Problems of Russian Academy of Sciences, Russia
Aleksey Buzmakov	Inria-LORIA (CNRS-Université de Lorraine), France
Artem Chernodub	Institute of Mathematical Machines and Systems Problems of Ukraine National Academy of Science, Ukraine

Vladimir Chernov	Institute for Image Processing of Russian Academy of Science, Russia
Ekaterina Chernyak	National Research University Higher School of Economics, Russia
Marina Chicheva	Samara National Research University, Russia
Bonaventura Coppola	IBM Research, USA
Hernani Costa	University of Malaga, Spain
Boris Dobrov	Lomonosov Moscow State University, Russia
Sofia Dokuka	National Research University Higher School of Economics, Russia
Florent Domenach	Akita International University, Japan
Alexey Drutsa	Lomonosov Moscow State University and Yandex, Russia
Julia Efremova	Eindhoven University of Technology, The Netherlands
Natalia Efremova	Moscow State Lomonosov University, Russia
Maria Eskevich	Radboud University Nijmegen, The Netherlands
Myasnikov Evgeny	Samara National Research University, Russia
Stefano Faralli	University of Mannheim, Germany
Victor Fedoseev	Image Processing Systems Institute of Russian Academy of Sciences and Samara National Research University, Russia
Michael Figurnov	Skolkovo Institute of Science and Technology, Russia
Elena Filatova	New York City College of Technology, USA
Mark Fishel	University of Tartu, Estonia
Thomas Francois	Université catholique de Louvain, Belgium
Oleksandr Frei	Schlumberger, Norway
Binyam Gebrekidan Gebre	Max Planck Institute for Psycholinguistics, The Netherlands
Natalia Grabar	STL CNRS Université Lille 3, France
Dmitry Granovsky	Yandex, Russia
Mena Habib	Maastricht University, The Netherlands
Dmitry Ilvovsky	National Research University Higher School of Economics, Russia
Vladimir Ivanov	Kazan Federal University, Russia
Dmitry Kan	SemanticAnalyzer, Finland
Nikolay Karpov	National Research University Higher School of Economics, Russia
Egor Kashkin	Vinogradov Russian Language Institute of Russian Academy of Sciences, Russia
Yury Katkov	Ecole Polytechnique Fédérale de Lausanne, Switzerland
Mehdi Kaytoue	LIRIS, INSA de Lyon, France
Alexander Kelmanov	Sobolev Institute of Mathematics of Siberian Branch of Russian Academy of Sciences, Russia
Andrey I. Kibzun	Moscow Aviation Institute, Russia
Victor Kitov	Lomonosov Moscow State University, Russia

Eduard Klyshinskii	Moscow State Institute of Electronics and Mathematics, Russia
Ekaterina Kochmar	University of Cambridge, UK
Sergei Koltcov	National Research University Higher School of Economics, Russia
Olessia Koltsova	National Research University Higher School of Economics, Russia
Alex Konduforov	AltexSoft, Ukraine
Natalia Konstantinova	University of Wolverhampton, UK
Andrey Kopylov	Tula State University, Russia
Kirill Korniyakov	Itseez and Lobachevsky State University of Nizhni Novgorod, Russia
Mikhail Korobov	ScrapingHub, Russia
Anton Korshunov	Institute for System Programming of Russian Academy of Sciences, Russia
Evgeny Kotelnikov	Vyatka State University, Russia
Olga Krasotkina	Tula State University, Russia
Tomas Krilavicius	Vytautas Magnus University, Lithuania
Valentina Kuskova	National Research University Higher School of Economics, Russia
Andrey Kutuzov	University of Oslo, Norway
Andrey Kuznetsov	Samara National Research University, Russia
Alexander Lepskiy	National Research University Higher School of Economics, Russia
Vadim Levit	Ariel University, Israel
Benjamin Lind	National Research University Higher School of Economics, Russia
Natalia Loukachevitch	Lomonosov Moscow State University, Russia
Olga Lyashevskaya	National Research University Higher School of Economics & Vinogradov Russian Language Institute of Russian Academy of Sciences, Russia
Ilya Markov	University of Amsterdam, The Netherlands
Luis Marujo	Carnegie Mellon University, USA and University of Lisbon, Portugal
Sérgio Matos	University of Aveiro, Portugal
Yelena Mejova	Qatar Computing Research Institute, Qatar
Benjamin Milde	Technische Universität Darmstadt, Germany
Olga Mitrofanova	St. Petersburg State University, Russia
Andrea Moro	Sapienza, Università di Roma, Italy
Hubert Naets	Université catholique de Louvain, Belgium
Vassilina Nikoulina	Xerox Research Center Europe, France
Damien Nouvel	Institut National des Langues et Civilisations Orientales, France
Dimitri Nowicki	Institute of Cybernetics of Ukraine National Academy of Science, Ukraine

Georgios Petasis	National Center for Scientific Research Demokritos, Greece
Andrey Philippovich	Bauman Moscow State Technical University, Moscow
Lidia Pivovarova	University of Helsinki, Finland
Vladimir Pleshko	RCO, Russia
Jonas Poelmans	Bigdatarobots, Cyprus
Alexander Porshnev	National Research University Higher School of Economics, Russia
Surya Prasath	University of Missouri-Columbia, USA
Delhibabu Radhakrishnan	Kazan Federal University, Russia
Carlos Ramisch	Aix Marseille University, France
Evgeniy Riabenko	Moscow Institute of Physics and Technology, Russia
Martin Riedl	Technische Universität Darmstadt, Germany
Alexandra Roshchina	Institute of Technology Tallaght, Ireland
Eugen Ruppert	Technische Universität Darmstadt, Germany
Mohammed Abdel-Mgeed M. Salem	Ain Shams University, Egypt
Grigory Sapunov	Stepic, Russia
Sheikh Muhammad Sarwar	University of Dhaka, Bangladesh and University of Massachusetts Amherst, USA
Andrey Savchenko	National Research University Higher School of Economics, Russia
Carsten Schnober	Deutsches Institut für Internationale Pädagogische Forschung, Germany
Marijn Schraagen	Utrecht University, The Netherlands
Alexander Semenov	National Research University Higher School of Economics, Russia
Oleg Seredin	Tula State University, Russia
Andrey Shcherbakov	Intel, Russia
Jan Snajder	University of Zagreb, Croatia
Evgeny Sokolov	Lomonosov Moscow State University, Germany
Christian Stab	Technische Universität Darmstadt, Germany
Vadim Strijov	Dorodnicyn Computing Centre of Russian Academy of Sciences, Russia
Maria Sukhareva	Goethe University of Frankfurt, Germany
Marina Suvorova	Lomonosov Moscow State University, Russia
Irina Temnikova	Qatar Computing Research Institute, Qatar
Christos Tryfonopoulos	University of the Peloponnese, Greece
Denis Turdakov	Institute for System Programming of Russian Academy of Sciences, Russia
Alexander Ulanov	HP Labs, Russia
Natalia Vassilieva	HP Labs, Russia
Dmitry Vetrov	Lomonosov Moscow State University, Russia
Ekaterina Vylomova	University of Melbourne, Australia

Patrick Watrin
Rostislav Yavorskiy

Université catholique de Louvain, Belgium
National Research University Higher School of
Economics, Russia

Seid Muhie Yimam
Marcos Zampieri
Olga Zvereva

Technische Universität Darmstadt, Germany
Saarland University, Germany
Ural Federal University, Russia

Additional Reviewers

Guillaume Bosc
Sujoy Chatterjee
Vladimir M. Chernov
Silvio Ricardo Cordeiro
Samuel Daylis
Anna Denisova
Svyatoslav Elizarov
Denis Fedyanin
Yuri Kan
Benjamin Milde

Alexander Minkin
Andrey Naumov
Paraskevi Raftopoulou
Oleg Slavin
Ivan Sterligov
Dmitry Ustalov
Natali Vaganova
Diliara Valeeva
Lefteris Zervakis

Sponsors

Golden sponsors

Exactpro
OK.Ru (Mail.Ru Group)

Bronze sponsor

IT Centre

Abstracts of Invited Talks

Analysis of Anisotropic Diffusion in Image Processing

Radhakrishnan Delhibabu

Kazan Federal University, Kazan, Russia
rdelhibabu@gmail.com

Abstract. Anisotropic diffusion is used widely in image processing for edge preserving filtering and image smoothing tasks. One of the important class of such model is by Perona and Malik [1], who used a gradient based diffusion to drive smoothing along edges and not across it. The contrast parameter used in the PM method needs to be carefully chosen to obtain optimal denoising results. Here we consider a local histogram based cumulative distribution approach for selecting this parameter in a data adaptive way so as to avoid manual tuning. We use spatial smoothing based diffusion coefficient along with adaptive contrast parameter estimation for obtaining better edge maps. Moreover, our experimental results indicate that this adaptive scheme performs well for a variety of noisy images. Thus, we obtain better peak signal to noise ratio and structural similarity scores with respect to fixed constant parameter values. The work proposes a modification of fuzzy diffusion coefficient that takes into account local pixel variability for better denoising and selective smoothing of edges.

Keywords: Image restoration · Anisotropic diffusion · Fuzzy edge detection · Contrast parameter · Local histogram · Diffusion coefficient · Denoising · Image processing

Acknowledgment. The speaker would like to thank Surya Prasath for the provided materials and help during the talk preparation.

References

1. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(7) (1990) 629–639
2. Canny, J.: A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **8**(6), 679–698 (1986)
3. Prasath, V.B.S.: A well-posed multiscale regularization scheme for digital image denoising. *Appl. Math. Comput. Sci.* **21**(4), 769–777 (2011)
4. Prasath, V.B.S., Delhibabu, R.: Image restoration with fuzzy coefficient driven anisotropic diffusion. In: *Swarm, Evolutionary, and Memetic Computing - 5th International Conference, SEMCCO 2014, Bhubaneswar, India, December 18–20, 2014*, 145–155 (2014). Revised Selected Papers

5. Prasath, V.B.S., Delhibabu, R.: Automatic contrast parameter estimation in anisotropic diffusion for image restoration. In: Analysis of Images, Social Networks and Texts - Third International Conference, AIST 2014, Yekaterinburg, Russia, April 10–12, 2014, 198–206 (2014). Revised Selected Papers
6. Prasath, V.B.S., Singh, A.: A hybrid convex variational model for image restoration. Appl. Math. Comput. **215**(10), 3655–3664 (2010)
7. Prasath, V.B.S., Singh, A.: Well-posed inhomogeneous nonlinear diffusion scheme for digital image denoising. J. Appl. Math. **2010**, 763847:1–763847:14 (2010)
8. Prasath, V.B.S., Vorotnikov, D.: On a system of adaptive coupled pdes for image restoration. J. Math. Imaging Vis. **48**(1), 35–52 (2014)

Finding Duplicate Labels in Behavioral Data: an Application for E-Sport Analytics

Mehdi Kaytoue

Université de Lyon, CNRS, INSA-Lyon, LIRIS UMR5205, 69621,
Villeurbanne, France
mehdi.kaytoue@insa-lyon.fr

Abstract. Analyzing behavioral data, by means of data mining, machine learning and visualization, helps answering several industrial challenges and proposing rich applications. When a trace is labeled by the user that generates it, models can be learned to accurately predict the user of an unknown trace in many domains (online security, target marketing, fraud detection) thanks to driving, typing, and mobility patterns. In online systems however, a user may have several virtual identities (avatar aliases): when this mapping is not known, the prediction accuracy drastically drops as the model considers each alias as a different user. In this invited talk, I will present several solutions to tackle this duplicate labels identification problem [1, 2]. We will consider online video games, and eSport in particular, as a use case. I will finish by presenting our recent results in this emerging domain, especially for the mining of strategic patterns from game behavioral data [3–6].

Keywords: Sport Analytics · eSport analytics · Behavioral data · Game data science · Video game · DOTA · StarCraft

References

1. Cavadenti, O., Codocedo, V., Boulicaut, J., Kaytoue, M.: When cyberathletes conceal their game: Clustering confusion matrices to identify avatar aliases. In: IEEE International Conference on Data Science and Advanced Analytics (DSAA) (2015)
2. Labernia, Q., Codocedo, V., Kaytoue, M., Robardet, C.: Découverte de labels dupliqués par l’exploration du treillis des classifieurs binaires. In: 16ème Journées Francophones Extraction et Gestion des Connaissances (EGC). Volume E-30 of RNTL, Hermann-Éditions, 255–266 (2016)
3. Kaytoue, M., Plantevit, M., Zimmermann, A., Bendimerad, A., Robardet, C.: Exceptional contextual subgraph mining. *Mach. Learn.* (2016)
4. Cavadenti, O., Codocedo, V., Boulicaut, J., Kaytoue, M.: What did i do wrong in my MOBA game? mining patterns discriminating deviant behaviours. In: IEEE International Conference on Data Science and Advanced Analytics (DSAA), 1–10 (2015)
5. Bosc, G., Tan, P., Boulicaut, J.F., Raïssi, C., Kaytoue, M.: A pattern mining approach to study strategy balance in RTS games. *IEEE Trans. Comput. Intell. AI Game* (99) (2015)
6. Low-Kam, C., Raïssi, C., Kaytoue, M., Pei, J.: Mining statistically significant sequential patterns. In: 13th International Conference on Data Mining (ICDM), IEEE Computer Society, pp. 488–497 (2013)

False Discovery Rate Control for Database Search Methods over Heterogeneous Biological Data

Attila Kertesz-Farkas

National Research University Higher School of Economics, Moscow, Russia
akerteszfarkas@hse.ru

Abstract. Database searching methods have become standard methods in identification of biological data, which often involves iteratively matching a query data against large protein sequence databases inexact manner. Accurate assignment is hindered by two problems. First, large number of hypotheses considered; i.e. a high score assigned to a match may not end up being statistically significant after multiple testing correction. Second, the hypothesis space is often combined by various types of data. Application of False Discovery Rate (FDR) control procedures over the combined hypothesis space in a single analysis can be dangerously leading to either overly conservative or overly liberal FDR within any subclasses. For these two problems, we propose a recently published new FDR control procedure, called Cascaded Search.

This method requires two inputs: (1) a user specified statistical confidence threshold, and (2) a series of peptide databases in a user specified order. For instance, such a cascade of databases could include fully tryptic, semitryptic, and nonenzymatic peptides or peptides with increasing numbers of modifications. Cascaded search then gradually searches the databases in a given order, sequestering at each stage query data that is identified with a specified statistical confidence. Our method has been compared to a standard procedure that lumps all of the peptides into a single database, as well as to a previously described Group-FDR procedure that computes the FDR separately within each database. We demonstrate, using simulated and real data, that cascaded search identifies more spectra at a fixed FDR threshold than that with either the ungrouped or grouped approach. Cascaded search thus provides a general method for maximizing the number of identified spectra in a statistically rigorous fashion and it is implemented in the open source Crux mass spectrometry analysis toolkit.

Keywords: Heterogeneous Biological Data • Inexact database searching • False Discovery Rate • Computational mass spectrometry

References

1. Kertesz-Farkas, A., Keich, U., Noble, W.S.: Tandem mass spectrum identification via cascaded search. *J. Proteome Res* **14**(8), 3027–3038 (2015). PMID: 26084232
2. McIlwain, S., Tamura, K., Kertesz-Farkas, A., Grant, C.E., Diament, B., Frewen, B., Howbert, J.J., Hoopmann, M.R., Kall, L., Eng, J.K., et al.: Crux: rapid open source protein tandem mass spectrometry analysis. *J. Proteome Res.* **13**(10), 448–4491 (2014)

Contents

Industry Talks

An Ontology-Driven Approach to Electronic Document Structure Design.	3
<i>Denis A. Nikiforov, Alexander B. Korchagin, and Ruslan L. Sivakov</i>	
Data Augmentation for Training of Noise Robust Acoustic Models.	17
<i>Tatiana Prisyach, Valentin Mendeleev, and Dmitry Ubskiy</i>	
Performance of Machine Learning Algorithms in Predicting Game Outcome from Drafts in Dota 2	26
<i>Aleksandr Semenov, Peter Romov, Sergey Korolev, Daniil Yashkov, and Kirill Neklyudov</i>	

Machine Learning and Data Analysis

Vote Aggregation Techniques in the Geo-Wiki Crowdsourcing Game: A Case Study	41
<i>Artem Baklanov, Steffen Fritz, Michael Khachay, Oleg Nurmukhametov, Carl Salk, Linda See, and Dmitry Shchepashchenko</i>	
On Complexity of Searching a Subset of Vectors with Shortest Average Under a Cardinality Restriction.	51
<i>Anton V. Ereemeev, Alexander V. Kel'manov, and Artem V. Pyatkin</i>	
The Problem of the Optimal Packing of the Equal Circles for Special Non-Euclidean Metric.	58
<i>Alexander L. Kazakov, Anna A. Lempert, and Huy L. Nguyen</i>	
Random Forest Based Approach for Concept Drift Handling	69
<i>Aleksei V. Zhukov, Denis N. Sidorov, and Aoife M. Foley</i>	

Social Networks

The Detailed Structure of Local Entrepreneurial Networks: Experimental Economic Study	81
<i>Dmitry B. Berg, Rustam H. Davletbaev, Yulia Y. Nazarova, and Olga M. Zvereva</i>	
Homophily Evolution in Online Networks: Who Is a Good Friend and When?.	91
<i>Sofia Dokuka, Diliara Valeeva, and Maria Yudkevich</i>	

The Structure of Organization: The Coauthorship Network Case	100
<i>Fedor Krasnov, Sofia Dokuka, and Rostislav Yavorskiy</i>	
Organizational Networks Revisited: Relational Predictors of Organizational Citizenship Behavior	108
<i>Valentina Kuskova, Elena Artyukhova, Rustam Kamalov, and Daria Danilova</i>	
Natural Language Processing	
Bigram Anchor Words Topic Model	121
<i>Arseniy Ashuha and Natalia Loukachevitch</i>	
Parallel Non-blocking Deterministic Algorithm for Online Topic Modeling . . .	132
<i>Oleksandr Frei and Murat Apishev</i>	
Flexible Context Extraction for Keywords in Russian Automatic Speech Recognition Results.	145
<i>Olga Khomitsevich, Kirill Boyarsky, Eugeny Kanevsky, Anna Bulusheva, and Valentin Mendelev</i>	
WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models	155
<i>Andrey Kutuzov and Elizaveta Kuzmenko</i>	
Morphological Analysis for Russian: Integration and Comparison of Taggers	162
<i>Elizaveta Kuzmenko</i>	
Anti-spoofing Methods for Automatic Speaker Verification System	172
<i>Galina Lavrentyeva, Sergey Novoselov, and Konstantin Simonchik</i>	
Combining Knowledge and CRF-Based Approach to Named Entity Recognition in Russian	185
<i>V.A. Mozharova and N.V. Loukachevitch</i>	
User Profiling in Text-Based Recommender Systems Based on Distributed Word Representations	196
<i>Anton Alekseev and Sergey Nikolenko</i>	
Constructing Aspect-Based Sentiment Lexicons with Topic Modeling	208
<i>Elena Tutubalina and Sergey Nikolenko</i>	
Human and Machine Judgements for Russian Semantic Relatedness	221
<i>Alexander Panchenko, Dmitry Ustalov, Nikolay Arefyev, Denis Paperno, Natalia Konstantinova, Natalia Loukachevitch, and Chris Biemann</i>	

Evaluating Distributional Semantic Models with Russian Noun-Adjective Compositions 236
Polina Panicheva, Ekaterina Protopopova, Grigoriy Bukia, and Olga Mitrofanova

Applying Word Embeddings to Leverage Knowledge Available in One Language in Order to Solve a Practical Text Classification Problem in Another Language 248
Andrew Smirnov and Valentin Mendelev

Analysis of Images and Video

Image Processing Algorithms with Structure Transferring Properties on the Basis of Gamma-Normal Model 257
Inessa Gracheva and Andrey Kopylov

Two Implementations of Probability Anomaly Detector Based on Different Vector Quantization Algorithms 269
Anna Denisova

Threefold Symmetry Detection in Hexagonal Images Based on Finite Eisenstein Fields 281
Alexander Karkishchenko and Valeriy Mnukhin

Reflection Symmetry of Shapes Based on Skeleton Primitive Chains. 293
Olesia Kushnir, Sofia Fedotova, Oleg Seredin, and Alexander Karkishchenko

Using Efficient Linear Local Features in the Copy-Move Forgery Detection Task 305
Andrey Kuznetsov and Vladislav Myasnikov

The Color Excitable Schrodinger Metamedium 314
Ekaterina Ostheimer, Valery Labunets, and Ivan Artemov

An Efficient Algorithm for Total Variation Denoising 326
Artyom Makovetskii, Sergei Voronin, and Vitaly Kober

Classification of Dangerous Situations for Small Sample Size Problem in Maintenance Decision Support Systems 338
Vladimir R. Milov and Andrey V. Savchenko

Ortho-Unitary Transforms, Wavelets and Splines 346
Ekaterina Ostheimer, Valery Labunets, and Ivan Artemov

A Real-Time Algorithm for Mobile Robot Mapping Based on
Rotation-Invariant Descriptors and Iterative Close Point Algorithm 357
A. Vokhmintcev and K. Yakovlev

Author Index 371