Lecture Notes in Computer Science 12440

More information about this series at http://www.springer.com/series/7409

Shin'ichi Satoh · Lucia Vadicamo ·
Arthur Zimek · Fabio Carrara ·
Ilaria Bartolini · Martin Aumüller ·
Björn Þór Jónsson · Rasmus Pagh (Eds.)

Similarity Search and Applications

13th International Conference, SISAP 2020
Copenhagen, Denmark, September 30 – October 2, 2020
Proceedings

Springer

Editors
Shin'ichi Satoh
National Institute of Informatics
Tokyo, Japan

Arthur Zimek
University of Southern Denmark
Odense M, Denmark

Ilaria Bartolini
University of Bologna
Bologna, Italy

Björn Þór Jónsson
IT University of Copenhagen
Copenhagen, Denmark

Lucia Vadicamo
ISTI-CNR
Pisa, Italy

Fabio Carrara
ISTI-CNR
Pisa, Italy

Martin Aumüller
IT University of Copenhagen
Copenhagen, Denmark

Rasmus Pagh
IT University of Copenhagen
Copenhagen, Denmark

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-030-60935-1 ISBN 978-3-030-60936-8 (eBook)
https://doi.org/10.1007/978-3-030-60936-8

LNCS Sublibrary: SL3 – Information Systems and Applications, incl. Internet/Web, and HCI

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume contains the papers presented at the 13th International Conference on Similarity Search and Applications (SISAP 2020), held during September 30 – October 2, 2020. The conference was planned to be hosted by the IT University of Copenhagen, Denmark. Due to the COVID-19 pandemic and international travel restrictions around the globe, however, SISAP 2020 had to be held as an online conference instead.

SISAP is an annual forum for researchers and application developers in the area of similarity data management. It focuses on the technological problems shared by numerous application domains, such as data mining, information retrieval, multimedia, computer vision, pattern recognition, computational biology, geography, biometrics, machine learning, and many others that make use of similarity search as a necessary supporting service.

From its roots as a regional workshop in metric indexing, SISAP has expanded to become the only international conference entirely devoted to the issues surrounding the theory, design, analysis, practice, and application of content-based and feature-based similarity search. The SISAP initiative has also created a repository (http://www.sisap.org/) serving the similarity search community, for the exchange of examples of real-world applications, source code for similarity indexes, and experimental testbeds and benchmark data sets. In addition, SISAP 2020 featured the 2020 edition of the SISAP Doctoral Symposium, for which a technical program was assembled, to give PhD students an opportunity to present their research ideas in an international research venue. The Doctoral Symposium indeed provided a forum that facilitated interactions among PhD students and stimulates feedback from more experienced researchers.

The call for papers welcomed full research papers, short research papers, as well as position and demonstration papers, with all manuscripts presenting previously unpublished research contributions.

We received 50 submissions from authors based in 22 different countries. The Program Committee (PC) was composed of 63 members from 26 countries. Each submission received at least three reviews, and the papers and reviews were thoroughly discussed by the chairs and PC members. Based on the reviews and discussions, the PC chairs accepted 19 full papers and 12 short papers (including 2 demonstration papers and 1 position paper), resulting in an acceptance rate of 38% for the full papers and 62% cumulative for full and short papers. After a separate review by the Doctoral Symposium Program Committee members, two Doctoral Symposium papers, giving a clear sample of emerging topics in similarity search and applications, were accepted for presentation and included in the program and proceedings.

The proceedings of SISAP are published by Springer as a volume in the *Lecture Notes in Computer Science* (LNCS) series. For SISAP 2020, as in previous years, extended versions of selected excellent papers were invited for publication in a special issue of the journal *Information Systems*. The conference also conferred a Best Paper

Award, a Best Student Paper Award, and a Best Doctoral Symposium Paper Award, as judged by the PC co-chairs and the Steering Committee.

Besides the presentations of the accepted papers, the conference program featured three keynote talks from outstanding scientists from industry and academia: Prof. Marcel Worring from University of Amsterdam, The Netherlands, Divesh Srivastava from AT&T Labs-Research, USA, and Ilya Razenshteyn from Microsoft Research, USA.

We would like to thank all the authors who submitted papers to SISAP 2020. We would also like to thank all members of the PC and the external reviewers for their effort and contribution to the conference. We want to extend our gratitude to the members of the Organizing Committee for the enormous amount of work they have done, and our sponsors and supporters for their generosity. Finally, we thank all the participants in the online event, who make up the thriving SISAP community.

September 2020

Shin'ichi Satoh
Lucia Vadicamo
Arthur Zimek
Fabio Carrara
Ilaria Bartolini
Martin Aumüller
Björn Þór Jónsson
Rasmus Pagh

Organization

General Chairs

Martin Aumüller	IT University of Copenhagen, Denmark
Björn Þór Jónsson	IT University of Copenhagen, Denmark
Rasmus Pagh	IT University of Copenhagen, Denmark

Program Committee Chairs

Shin'ichi Satoh	National Institute of Informatics, Japan
Lucia Vadicamo	ISTI-CNR, Italy
Arthur Zimek	University of Southern Denmark, Denmark

Doctoral Symposium Program Committee Chair

Ilaria Bartolini	University of Bologna, Italy

Publication Chair

Fabio Carrara	ISTI-CNR, Italy

Steering Committee

Laurent Amsaleg	CNRS-IRISA, France
Edgar Chávez	CICESE, Mexico
Michael E. Houle	National Institute of Informatics, Japan
Pavel Zezula	Masaryk University, Czech Republic

Program Committee

Giuseppe Amato	ISTI-CNR, Italy
Laurent Amsaleg	CNRS-IRISA, France
Fabrizio Angiulli	University of Calabria, Italy
James Bailey	The University of Melbourne, Australia
Christian Beecks	University of Münster, Germany
Virendra Bhavsar	University of New Brunswick, Canada
Panagiotis Bouros	Johannes Gutenberg University Mainz, Germany
Benjamin Bustos	University of Chile, Chile
Selçuk Candan	Arizona State University, USA
Aniket Chakrabarti	Microsoft AI & Research, India
Edgar Chávez	CICESE, Mexico
Richard Chbeir	University Pau and Pays de l'Adour, France

Richard Connor	University of St Andrews, UK
Petros Daras	Information Technologies Institute, Greece
Alan Dearle	University of St Andrews, UK
Vlastislav Dohnal	Masaryk University, Czech Republic
Vladimir Estivill-Castro	Griffith University, Australia
Andrea Esuli	ISTI-CNR, Italy
Rolf Fagerberg	University of Southern Denmark, Denmark
Fabrizio Falchi	ISTI-CNR, Italy
Claudio Gennaro	ISTI-CNR, Italy
Magnus Lie Hetland	Norwegian University of Science and Technology, Norway
Thi Thao Nguyen Ho	Aalborg University, Denmark
Michael E. Houle	National Institute of Informatics, Japan
Ichiro Ide	Nagoya University, Japan
Kyoung-Sook Kim	National Institute of Advanced Industrial Science and Technology, Japan
Peer Kröger	Ludwig Maximilians University of Munich, Germany
Jakub Lokoč	Charles University, Czech Republic
Rui Mao	Shenzhen University, China
Stephane Marchand-Maillet	University of Geneva, Switzerland
Yusuke Matsui	The University of Tokyo, Japan
Luisa Micó	University of Alicante, Spain
Henning Müller	HES-SO, Switzerland
Chong-Wah Ngo	City University of Hong Kong, Hong Kong
Vincent Oria	New Jersey Institute of Technology, USA
Deepak P.	Queen's University Belfast, UK
Rodrigo Paredes	University of Talca, Chile
Marco Patella	University of Bologna, Italy
Oscar Pedreira	University of A Coruña, Spain
Raffaele Perego	ISTI-CNR, Italy
Miloš Radovanović	University of Novi Sad, Serbia
Nora Reyes	Universidad Nacional de San Luis, Argentina
Marcela Xavier Ribeiro	Federal University of São Carlos, Brazil
Kunihiko Sadakane	The University of Tokyo, Japan
Maria Luisa Sapino	Università di Torino, Italy
Erich Schubert	Technical University of Dortmund, Germany
Matthias Schubert	Ludwig Maximilians University of Munich, Germany
Thomas Seidl	Ludwig Maximilians University of Munich, Germany
Tetsuo Shibuya	The University of Tokyo, Japan
Tomas Skopal	Charles University, Czech Republic
Yasuo Tabei	RIKEN Center for Advanced Intelligence Project, Japan
Joe Tekli	Lebanese American University, Lebanon
Nenad Tomasev	DeepMind, UK
Agma J. M. Traina	University of São Paulo, Brazil
Caetano Traina	University of São Paulo, Brazil

Goce Trajcevski	Iowa State University, USA
Takashi Washio	Osaka University, Japan
Marcel Worring	University of Amsterdam, The Netherlands
Kaoru Yoshida	Sony Computer Science Laboratories, Inc., Japan
Pavel Zezula	Masaryk University, Czech Republic
Kaiping Zheng	National University of Singapore, Singapore
Zhi-Hua Zhou	Nanjing University, China
Andreas Züfle	George Mason University, USA

Doctoral Symposium Program Committee

Selçuk Candan	Arizona State University, USA
Pavel Zezula	Masaryk University, Czech Republic

Special Session Organizers

Giuseppe Amato	ISTI-CNR, Italy
Laurent Amsaleg	CNRS-IRISA, France
Fabio Carrara	ISTI-CNR, Italy
Fabrizio Falchi	ISTI-CNR, Italy
Claudio Gennaro	ISTI-CNR, Italy
Michael E. Houle	National Institute of Informatics, Japan
Sanjiv Kumar	Google Research, USA
Rasmus Pagh	IT University of Copenhagen, Denmark
Anshumali Shrivastava	Rice University, USA

Additional Reviewers

Anna Beer
Fabian Berns
Felix Borutta
Fabio Carrara
Mirela Teixeira Cazzolato
Fabio Fassetti
Luca Ferragina
Tahrima Hashem
Daniyal Kazempour
Oscar Cuadros Linares
Gabriele Lagani
Elio Mansour
Vladimir Mic
Alejandro Moreo Fernández
Ladislav Peska
Phil Sun
Erik Thordsen
Zhaozhuo Xu

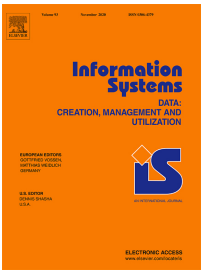
Sponsors

IT University of Copenhagen

Google

Springer

Information Systems, Elsevier



Abstracts of Keynotes

Interactive Exploration using Hypergraphs

Marcel Worring

University of Amsterdam, The Netherlands

Abstract. Interactive exploration of a multimedia collection, ranging from search to browsing, requires various tasks to be supported by the system. Categorization, in which each item receives a membership score, provides a unifying framework for many of these tasks that can now, with specialized efficient high-dimensional indexing, interactively be performed even for very large collections. It also provides a proper basis for the notoriously difficult task of evaluating interactive exploration. Categorization is primarily based on the learned features of the items in the collection, possibly implicitly supported by metric learning. It does not explicitly capture the similarity or knowledge-based relations among items in the collection. Hypergraphs generalize graphs by having edges which can connect any number of nodes instead of just two. In doing so they are effectively combining categories and similarity-based relations in one model. Recent advances in graph-convolutional networks bring new opportunities to learning using hypergraphs, predicting a hyperedge membership score that captures both similarity among the elements as well as group membership. In this talk, we highlight progress made in hypergraph learning and how it leads to new opportunities for interactive exploration of multimedia content.

Exploiting Similarity Relationships to Repair Graphs

Divesh Srivastava

AT&T Labs-Research, USA

Abstract. Graphs are a flexible way to represent data in a variety of applications, with nodes representing domain-specific entities (e.g., records in entity resolution, products categories in a taxonomy) and edges capturing a variety of relationships between these entities (e.g., a linkage relationship between records in entity resolution, a category-subcategory relationship between product categories in a taxonomy). Often, the edges in this graph are inferred based on similarity relationships between nodes and are noisy, in that some edges are missing (i.e., real-world relationships that do not have corresponding edges in the graph) and some edges are spurious (i.e., edges in the graph that do not have corresponding real-world relationships). Directly analyzing such graphs can lead to undesirable outcomes, making it important to repair noisy graphs. In this talk, we describe an approach that takes advantage of properties of real-world relationships and their estimated probabilities to ask oracle queries (an abstraction of crowdsourcing) to efficiently repair the noisy graphs. We illustrate this approach for the case of graphs that are unions of cliques (which is the case for entity resolution) and graphs that are tree-structured (which is the case for taxonomies), and present theoretical and empirical results for these cases.

Scalable Nearest Neighbor Search for Optimal Transport

Ilya Razenshteyn

Microsoft Research, USA

Abstract. The Optimal Transport (aka Wasserstein) distance is an increasingly popular similarity measure for structured data domains, such as images or text documents. This raises the necessity for fast nearest neighbor search with respect to this distance, a problem that poses a substantial computational bottleneck for various tasks on massive datasets. In this talk, I will discuss fast tree-based approximation algorithms for searching nearest neighbors with respect to the Wasserstein-1 distance. I will start with describing a standard tree-based technique, known as QuadTree, which has been previously shown to obtain good results. Then I'll introduce a variant of this algorithm, called FlowTree, and show that it achieves better accuracy, both in theory and in practice. In particular, the accuracy of FlowTree is in line with previous high-accuracy methods, while its running time is much faster. The talk is based on a joint work with Arturs Backurs, Yihe Dong, Piotr Indyk, and Tal Wagner. The paper[1] and code[2] is available.

[1] https://arxiv.org/abs/1910.04126.

[2] https://github.com/ilyaraz/ot_estimators.

Contents

Artificial Intelligence and Similarity

Demo and Position Papers

Doctoral Symposium