Lecture Notes in Computer Science 12391

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

Sven Hartmann · Josef Küng ·
Gabriele Kotsis · A Min Tjoa ·
Ismail Khalil (Eds.)

Database and Expert Systems Applications

31st International Conference, DEXA 2020
Bratislava, Slovakia, September 14–17, 2020
Proceedings, Part I

Springer

Editors
Sven Hartmann
Clausthal University of Technology
Clausthal-Zellerfeld, Germany

Gabriele Kotsis
Johannes Kepler University of Linz
Linz, Austria

Ismail Khalil
Johannes Kepler University of Linz
Linz, Austria

Josef Küng
Johannes Kepler University of Linz
Linz, Austria

A Min Tjoa
IFS
Vienna University of Technology
Vienna, Wien, Austria

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-030-59002-4 ISBN 978-3-030-59003-1 (eBook)
https://doi.org/10.1007/978-3-030-59003-1

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 31st International Conference on Database and Expert Systems Applications (DEXA 2020). This year, DEXA was held as a virtual conference during September 14–17, 2020, instead of as it was originally planned to be held in Bratislava, Slovakia.

On behalf of the Program Committee we commend these papers to you and hope you find them useful.

Database, information, and knowledge systems have always been a core subject of computer science. The ever increasing need to distribute, exchange, and integrate data, information, and knowledge has added further importance to this subject. Advances in the field will help facilitate new avenues of communication, to proliferate interdisciplinary discovery, and to drive innovation and commercial opportunity.

DEXA is an international conference series which showcases state-of-the-art research activities in database, information, and knowledge systems. The conference and its associated workshops provide a premier annual forum to present original research results and to examine advanced applications in the field. The goal is to bring together developers, scientists, and users to extensively discuss requirements, challenges, and solutions in database, information, and knowledge systems.

DEXA 2020 solicited original contributions dealing with all aspects of database, information, and knowledge systems. Suggested topics included, but were not limited to:

- Acquisition, Modeling, Management and Processing of Knowledge
- Authenticity, Privacy, Security, and Trust
- Availability, Reliability and Fault Tolerance
- Big Data Management and Analytics
- Consistency, Integrity, Quality of Data
- Constraint Modeling and Processing
- Cloud Computing and Database-as-a-Service
- Database Federation and Integration, Interoperability, Multi-Databases
- Data and Information Networks
- Data and Information Semantics
- Data Integration, Metadata Management, and Interoperability
- Data Structures and Data Management Algorithms
- Database and Information System Architecture and Performance
- Data Streams and Sensor Data
- Data Warehousing
- Decision Support Systems and Their Applications
- Dependability, Reliability, and Fault Tolerance
- Digital Libraries and Multimedia Databases
- Distributed, Parallel, P2P, Grid, and Cloud Databases
- Graph Databases

- Incomplete and Uncertain Data
- Information Retrieval
- Information and Database Systems and Their Applications
- Mobile, Pervasive, and Ubiquitous Data
- Modeling, Automation, and Optimization of Processes
- NoSQL and NewSQL Databases
- Object, Object-Relational, and Deductive Databases
- Provenance of Data and Information
- Semantic Web and Ontologies
- Social Networks, Social Web, Graph, and Personal Information Management
- Statistical and Scientific Databases
- Temporal, Spatial, and High Dimensional Databases
- Query Processing and Transaction Management
- User Interfaces to Databases and Information Systems
- Visual Data Analytics, Data Mining, and Knowledge Discovery
- WWW, Databases and Web Services
- Workflow Management and Databases
- XML and Semi-structured Data

Following the call for papers which yielded 190 submissions, there was a rigorous refereeing process that saw each submission reviewed by three to five international experts. The 38 submissions judged best by the Program Committee were accepted as full research papers, yielding an acceptance rate of 20%. A further 20 submissions were accepted as short research papers.

As is the tradition of DEXA, all accepted papers are published by Springer. Authors of selected papers presented at the conference were invited to submit substantially extended versions of their conference papers for publication in special issues of international journals. The submitted extended versions underwent a further review process.

We wish to thank all authors who submitted papers and all conference participants for the fruitful discussions.

This year we have five keynote talks addressing emerging trends in the database and artificial intelligence community:

- “Knowledge Graphs for Drug Discovery” by Prof. Ying Ding (The University of Texas at Austin, USA)
- “Incremental Learning and Learning with Drift” by Prof. Barbara Hammer (CITEC Centre of Excellence, Bielefeld University, Germany)
- “From Sensors to Dempster-Shafer Theory and Back: the Axiom of Ambiguous Sensor Correctness and its Applications” by Prof. Dirk Draheim (Tallinn University of Technology, Estonia)
- “Knowledge Availability and Information Literacies” by Dr. Gerald Weber (The University of Auckland, New Zealand)
- “Explainable Fact Checking for Statistical and Property Claims” by Paolo Papotti (EURECOM, France)

In addition, we had a panel discussion on "The Age of Science-making Machines" led by Prof. Stéphane Bressan (National University of Singapore, Singapore).

This edition of DEXA features three international workshops covering a variety of specialized topics:

- BIOKDD 2020: 11th International Workshop on Biological Knowledge Discovery from Data
- IWCFS 2020: 4th International Workshop on Cyber-Security and Functional Safety in Cyber-Physical Systems
- MLKgraphs 2020: Second International Workshop on Machine Learning and Knowledge Graphs

The success of DEXA 2020 is a result of collegial teamwork from many individuals. We like to thank the members of the Program Committee and the external referees for their timely expertise in carefully reviewing the submissions.

Warm thanks to Ismail Khalil and the conference organizers as well as all workshop organizers.

We would also like to express our thanks to all institutions actively supporting this event, namely:

- Comenius University Bratislava (who was prepared to host the conference)
- Institute of Telekoopertion, Johannes Kepler University Linz (JKU)
- Software Competence Center Hagenberg (SCCH)
- International Organization for Information Integration and Web based Applications and Services (@WAS)

We hope you enjoyed the conference program.

September 2020

Sven Hartmann
Josef Küng

Organization

Program Committee Chairs

Sven Hartmann	Clausthal University of Technology, Germany
Josef Küng	Johannes Kepler University Linz, Austria

Steering Committee

Gabriele Kotsis	Johannes Kepler University Linz, Austria
A Min Tjoa	Vienna University of Technology, Austria
Ismail Khalil	Johannes Kepler University Linz, Austria

Program Committee and Reviewers

Javier Nieves Acedo	Azterlan, Spain
Sonali Agarwal	IIIT, India
Riccardo Albertoni	CNR-IMATI, Italy
Idir Amine Amarouche	USTHB, Algeria
Rachid Anane	Coventry University, UK
Mustafa Atay	Winston-Salem State University, USA
Faten Atigui	CEDRIC, CNAM, France
Ladjel Bellatreche	LIAS, ENSMA, France
Nadia Bennani	LIRIS, INSA de Lyon, France
Karim Benouaret	Université de Lyon, France
Djamal Benslimane	Université de Lyon, France
Morad Benyoucef	University of Ottawa, Canada
Catherine Berrut	LIG, Université Joseph Fourier Grenoble I, France
Vasudha Bhatnagar	University of Delhi, India
Athman Bouguettaya	The University of Sydney, Australia
Omar Boussai	ERIC Laboratory, France
Stephane Bressan	National University of Singapore, Singapore
Sharma Chakravarthy	The University of Texas at Arlington, USA
Cindy Chen	University of Massachusetts Lowell, USA
Gang Chen	Victoria University of Wellington, New Zealand
Max Chevalier	IRIT, France
Soon Ae Chun	City University of New York, USA
Alfredo Cuzzocrea	University of Calabria, Italy
Debora Dahl	Conversational Technologies, USA
Jérôme Darmont	Université de Lyon, France
Soumyava Das	Teradata Labs, USA
Vincenzo Deufemia	University of Salerno, Italy
Dejing Dou	University of Oregon, USA

Cedric Du Mouza	CNAM, France
Umut Durak	German Aerospace Center (DLR), Germany
Suzanne Embury	The University of Manchester, UK
Markus Endres	University of Passau, Germany
Nora Faci	Université de Lyon, France
Bettina Fazzinga	ICAR-CNR, Italy
Stefano Ferilli	Università di Bari, Italy
Flavio Ferrarotti	Software Competence Centre Hagenberg, Austria
Flavius Frasincar	Erasmus University Rotterdam, The Netherlands
Bernhard Freudenthaler	Software Competence Center Hagenberg, Austria
Steven Furnell	Plymouth University, UK
Aryya Gangopadhyay	University of Maryland Baltimore County, USA
Manolis Gergatsoulis	Ionian University, Greece
Javad Ghofrani	HTW Dresden University of Applied Sciences, Germany
Vikram Goyal	IIIT Delhi, India
Sven Groppe	University of Lübeck, Germany
William Grosky	University of Michigan, USA
Francesco Guerra	Università di Modena e Reggio Emilia, Italy
Giovanna Guerrini	DISI, University of Genova, Italy
Allel Hadjali	LIAS, ENSMA, France
Abdelkader Hameurlain	IRIT, Paul Sabatier University, France
Takahiro Hara	Osaka University, Japan
Ionut Iacob	Georgia Southern University, USA
Hamidah Ibrahim	Universiti Putra Malaysia, Malaysia
Sergio Ilarri	University of Zaragoza, Spain
Abdessamad Imine	Loria, France
Peiquan Jin	University of Science and Technology of China, China
Anne Kao	Boeing, USA
Anne Kayem	Hasso Plattner Institute, University of Potsdam, Germany
Uday Kiran	The University of Tokyo, Japan
Carsten Kleiner	University of Applied Science and Arts Hannover, Germany
Henning Koehler	Massey University, New Zealand
Michal Kratky	VSB-Technical University of Ostrava, Czech Republic
Petr Kremen	Czech Technical University in Prague, Czech Republic
Anne Laurent	LIRMM, University of Montpellier, France
Lenka Lhotska	Czech Technical University in Prague, Czech Republic
Wenxin Liang	Chongqing University of Posts and Telecommunications, China
Chuan-Ming Liu	National Taipei University of Technology, Taiwan
Hong-Cheu Liu	School of Information Technology and Mathematical Sciences, Australia
Jorge Lloret	University of Zaragoza, Spain
Hui Ma	Victoria University of Wellington, New Zealand

Qiang Ma	Kyoto University, Japan
Zakaria Maamar	Zayed University, UAE
Sanjay Madria	Missouri S&T, USA
Jorge Martinez-Gil	Software Competence Center Hagenberg, Austria
Elio Masciari	Federico II University, Italy
Atif Mashkoor	Software Competence Center Hagenberg, Austria
Jun Miyazaki	Tokyo Institute of Technology, Japan
Lars Moench	University of Hagen, Germany
Riad Mokadem	Paul Sabatier University, Pyramide Team, France
Anirban Mondal	The University of Tokyo, Japan
Yang-Sae Moon	Kangwon National University, South Korea
Franck Morvan	IRIT, Paul Sabatier University, France
Francesco D. Muñoz-Escoí	Universitat Politècnica de València, Spain
Ismael Navas-Delgado	University of Malaga, Spain
Wilfred Ng	The Hong Kong University of Science and Technology, Hong Kong
Teste Olivier	IRIT, France
Carlos Ordonez	University of Houston, USA
Marcin Paprzycki	Systems Research Institute, Polish Academy of Sciences, Poland
Oscar Pastor Lopez	Universitat Politècnica de València, Spain
Clara Pizzuti	CNR-ICAR, Italy
Elaheh Pourabbas	CNR, Italy
Birgit Proell	Johannes Kepler University Linz, Austria
Rodolfo Resende	Federal University of Minas Gerais, Brazil
Werner Retschitzegger	Johannes Kepler University Linz, Austria
Claudia Roncancio	Grenoble Alps University, France
Viera Rozinajova	Slovak University of Technology in Bratislava, Slovakia
Massimo Ruffolo	ICAR-CNR, Italy
Shelly Sachdeva	National Institute of Technology Delhi, India
Marinette Savonnet	University of Burgundy, France
Stefanie Scherzinger	Universität Passau, German
Wieland Schwinger	Johannes Kepler University Linz, Austria
Florence Sedes	IRIT, Paul Sabatier University, France
Nazha Selmaoui-Folcher	University of New Caledonia, New Caledonia
Michael Sheng	Macquarie University, Australia
Patrick Siarry	Université de Paris 12, France
Gheorghe Cosmin Silaghi	Babes-Bolyai University, Romania
Hala Skaf-Molli	University of Nantes, LS2N, France
Bala Srinivasan	Monash University, Australia
Stephanie Teufel	University of Fribourg, iimt, Switzerland
Jean-Marc Thévenin	IRIT, Université Toulouse I, France
Vicenc Torra	University Skövde, Sweden
Traian Marius Truta	Northern Kentucky University, USA
Lucia Vaira	University of Salento, Italy

Ismini Vasileiou	De Montfort University, UK
Krishnamurthy Vidyasankar	Memorial University, Canada
Marco Vieira	University of Coimbra, Portugal
Piotr Wisniewski	Nicolaus Copernicus University, Poland
Ming Hour Yang	Chung Yuan Chritian University, Taiwan
Xiaochun Yang	Northeastern University, China
Haruo Yokota	Tokyo Institute of Technology, Japan
Qiang Zhu	University of Michigan, USA
Yan Zhu	Southwest Jiaotong University, China
Marcin Zimniak	Leipzig University, Germany
Ester Zumpano	University of Calabria, Italy

Organizers

scch
software competence center
hagenberg

JKU
Institute for
Telecooperation

International Organization for
@WAS
www.iiwas.org
Information Integration and
Web-based Applications & Services

Abstracts of Keynote Talks

Knowledge Graph for Drug Discovery

Ying Ding

The University of Texas at Austin, USA

Abstract. A critical barrier in current drug discovery is the inability to utilize public datasets in an integrated fashion to fully understand the actions of drugs and chemical compounds on biological systems. There is a need to intelligently integrate heterogeneous datasets pertaining to compounds, drugs, targets, genes, diseases, and drug side effects now available to enable effective network data mining algorithms to extract important biological relationships. In this talk, we demonstrate the semantic integration of 25 different databases and showcase the cutting-edge machine learning and deep learning algorithms to mine knowledge graphs for deep insights, especially the latest graph embedding algorithm that outperforms baseline methods for drug and protein binding predictions.

Incremental Learning and Learning with Drift

Barbara Hammer

CITEC Centre of Excellence, Bielefeld University, Germany

Abstract. Neural networks have revolutionized domains such as computer vision or language processing, and learning technology is included in everyday consumer products. Yet, practical problems often render learning surprisingly difficult, since some of the fundamental assumptions of the success of deep learning are violated. As an example, only few data might be available for tasks such as model personalization, hence few shot learning is required. Learning might take place in non-stationary environments such that models face the stability-plasticity dilemma. In such cases, applicants might be tempted to use models for settings they are not intended for, such that invalid results are unavoidable.

Within the talk, I will address three challenges of machine learning when dealing with incremental learning tasks, addressing the questions: how to learn reliably given few examples only, how to learn incrementally in non-stationary environments where drift might occur, and how to enhance machine learning models by an explicit reject option, such that they can abstain from classification if the decision is unclear

From Sensors to Dempster-Shafer Theory and Back: The Axiom of Ambiguous Sensor Correctness and Its Applications

Dirk Draheim

Tallinn University of Technology, Estonia
dirk.draheim@taltech.ee

Abstract. Since its introduction in the 1960s, Dempster-Shafer theory became one of the leading strands of research in artificial intelligence with a wide range of applications in business, finance, engineering, and medical diagnosis. In this paper, we aim to grasp the essence of Dempster-Shafer theory by distinguishing between ambiguous-and-questionable and ambiguous-but-correct perceptions. Throughout the paper, we reflect our analysis in terms of signals and sensors as a natural field of application. We model ambiguous-and-questionable perceptions as a probability space with a quantity random variable and an additional perception random variable (Dempster model). We introduce a correctness property for perceptions. We use this property as an axiom for ambiguous-but-correct perceptions. In our axiomatization, Dempster's lower and upper probabilities do not have to be postulated: they are consequences of the perception correctness property. Furthermore, we outline how Dempster's lower and upper probabilities can be understood as best possible estimates of quantity probabilities. Finally, we define a natural knowledge fusion operator for perceptions and compare it with Dempster's rule of combination.

Knowledge Availability and Information Literacies

Gerald Weber

The University of Auckland, New Zealand

Abstract. At least since Tim Berners-Lee's call for 'Raw Data Now' in 2009, which he combined with a push for linked data as well, the question has been raised how to make the wealth of data and knowledge available to the citizens of the world. We will set out to explore the many facets and multiple layers of this problem, leading up to the question of how we as users will access and utilize the knowledge that should be available to us.

Explainable Fact Checking for Statistical and Property Claims

Paolo Papotti

EURECOM, France

Abstract. Misinformation is an important problem but fact checkers are overwhelmed by the amount of false content that is produced online every day. To support fact checkers in their efforts, we are creating data-driven verification methods that use structured datasets to assess claims and explain their decisions. For statistical claims, we translate text claims into SQL queries on relational databases. We exploit text classifiers to propose validation queries to the users and rely on tentative execution of query candidates to narrow down the set of alternatives. The verification process is controlled by a cost-based optimizer that considers expected verification overheads and the expected claim utility as training samples. For property claims, we use the rich semantics in knowledge graphs (KGs) to verify claims and produce explanations. As information in a KG is inevitably incomplete, we rely on rule discovery and on text mining to gather the evidence to assess claims. Uncertain rules and facts are turned into logical programs and the checking task is modeled as a probabilistic inference problem. Experiments show that both methods enable the efficient and effective labeling of claims with interpretable explanations, both in simulations and in real world user studies with 50% decrease in verification time. Our algorithms are demonstrated in a fact checking website (https://coronacheck.eurecom.fr), which has been used by more than twelve thousand users to verify claims related to the coronavirus disease (COVID-19) spread and effect.

“How Many Apples?” or the Age of Science-Making Machines (Panel)

Stéphane Bressan (Panel Chair)

National University of Singapore, Singapore

Abstract. Isaac Newton most likely did not spend much time observing apples falling from trees. Galileo Galilei probably never threw anything from the tower of Pisa. They conducted thought experiments.

What can big data, data science, and artificial intelligence contribute to the creation of scientific knowledge? How can advances in computing, communication, and control improve or positively disrupt the scientific method?

Richard Feynman once explained the scientific method as follows. “In general, we look for a new law by the following process. First, we guess it; don’t laugh that is really true. Then we compute the consequences of the guess to see what, if this is right, if this law that we guessed is right, we see what it would imply. And then we compare those computation results to nature, or, we say, compare to experiment or experience, compare directly with observations to see if it works. If it disagrees with experiment, it’s wrong and that simple statement is the key to science.” He added euphemistically that “It is therefore not unscientific to take a guess.”

Can machines help create science?

The numerous advances of the many omics constitute an undeniable body of evidence that computing, communication, and control technologies, in the form of high-performance computing hardware, programming frameworks, algorithms, communication networks, as well as storage, sensing, and actuating devices, help scientists and make the scientific process significantly more efficient and more effective. Everyone acknowledges the unmatched ability of machines to streamline measurements and to process large volumes of results, to facilitate complex modeling, and to run complex computations and extensive simulations. The only remaining question seems to be the extent of their unexplored potential.

Furthermore, the media routinely report new spectacular successes of big data analytics and artificial intelligence that suggest new opportunities. Scientists are discussing physics-inspired machine learning. We are even contemplating the prospect of breaking combinatorial barriers with quantum computers. However, except, possibly for the latter, one way or another, it all seems about heavy-duty muscle-flexing without much subtlety nor finesse.

Can machines take a guess?

Although the thought processes leading to the guesses from which theories are built are laden with ontological, epistemological, and antecedent theoretical assumptions, and the very formulation of the guesses assumes certain conceptual views, scientists seem to have been able to break through those glass ceilings again and again and invent entirely new concepts. Surely parallel computing, optimization algorithms, reinforcement learning, or genetic algorithms can assist

in the exploration of the space of combinatorial compositions of existing concepts. In the words of Feynman again: "We set up a machine, a great computing machine, which has a random wheel and if it makes a succession of guesses and each time it guesses a hypothesis about how nature should work computes immediately the consequences and makes a comparison to a list of experimental results that it has at the other hand. In other words, guessing is a dumb man's job. Actually, it is quite the opposite and I will try to explain why." He continues: "The problem is not to make, to change or to say that something might be wrong but to replace it with something and that is not so easy."

Can machines create new concepts?

The panelists are asked to share illustrative professional experiences, anecdotes, and thoughts, as well as their enthusiasm and concerns, regarding the actuality and potential of advances in computing, communication, and control in improving and positively disrupting the scientific process.

Contents – Part I

Contents – Part II

Knowledge Discovery

Machine Learning

Semantic Web and Ontologies