Lecture Notes in Computer Science 12390

Founding Editors

Gerhard Goos
Karlsruhe Institute of Technology, Karlsruhe, Germany
Juris Hartmanis
Cornell University, Ithaca, NY, USA

Editorial Board Members

Elisa Bertino
Purdue University, West Lafayette, IN, USA
Wen Gao
Peking University, Beijing, China
Bernhard Steffen
TU Dortmund University, Dortmund, Germany
Gerhard Woeginger
RWTH Aachen, Aachen, Germany
Moti Yung
Columbia University, New York, NY, USA

More information about this series at http://www.springer.com/series/8637

Abdelkader Hameurlain ·
A Min Tjoa · Richard Chbeir ·
Yannis Manolopoulos · Hiroshi Ishikawa ·
Sergio Ilarri · Apostolos Papadopoulos (Eds.)

Transactions on Large-Scale Data- and Knowledge-Centered Systems XLV

Special Issue on Data Management and Knowledge Extraction in Digital Ecosystems

Springer

Editors-in-Chief
Abdelkader Hameurlain
IRIT, Paul Sabatier University
Toulouse, France

Guest Editors
Richard Chbeir
University of Pau and Pays de l'Adour
Anglet, France

Hiroshi Ishikawa
Tokyo Metropolitan University
Tokyo, Japan

Apostolos Papadopoulos
Aristotle University of Thessaloniki
Thessaloniki, Greece

A Min Tjoa
Vienna University of Technology
Vienna, Austria

Yannis Manolopoulos
Open University of Cyprus
Nicosia, Cyprus

Sergio Ilarri
University of Zaragoza
Zaragoza, Spain

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISSN 1869-1994 ISSN 2510-4942 (electronic)
Transactions on Large-Scale Data- and Knowledge-Centered Systems
ISBN 978-3-662-62307-7 ISBN 978-3-662-62308-4 (eBook)
https://doi.org/10.1007/978-3-662-62308-4

© Springer-Verlag GmbH Germany, part of Springer Nature 2020
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, expressed 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.

This Springer imprint is published by the registered company Springer-Verlag GmbH, DE part of Springer Nature.
The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

Preface

In the world of the Internet of Things (IoT), the recent rapid growth and use of connected objects leads to the emergence of virtual environments composed of multiple and independent entities such as individuals, organizations, services, software, and applications sharing one or several missions and focusing on the interactions and interrelationships among them. These digital ecosystems exhibit self-organizing environments where the underlying resources mainly comprehend data management and computational collective intelligence. Due to the multidisciplinary nature of digital ecosystems and their characteristics, it is highly complex to provide a poor understanding as to how managing data will empower digital ecosystems to be innovative and value-creating. The application of Information Technologies has the potential to enable the understanding of how entities request to create benefits and added values, impacting business practices and knowledge. This context introduces many new challenges from different theoretical and practical points of view.

This special issue aims to assess the current status and technologies, as well as to outline the major challenges and future perspectives, related to the data management of digital ecosystems. It includes eight papers that were selected after a very tight peer review, in which each paper was reviewed by three reviewers. Several topics are addressed in this special issue, but mainly: data analysis, information extraction, blockchains, and big data. It is organized as follows.

In the first paper of this special issue, Demetris Trihinas proposes "Interoperable Data Extraction and Analytics Queries over Blockchains." Here, the author explores the explosion of interests by diverse organizations for deploying their services on blockchains to exploit decentralized transaction governance and advanced cryptographic protocols, fostering the emergence of new challenges for distributed ledger technologies (DLTs). The next generation of blockchain services are now extending well beyond cryptocurrencies, accumulating and storing vast amounts of data. Therefore, the need to efficiently extract data over blockchains and subsequently foster data analytics, is more evident than ever. However, despite the wide public interest and the release of several frameworks, efficiently accessing and processing data from blockchains still imposes significant challenges. First, the paper introduces the key limitations faced by organizations in need for efficiently accessing and managing data over DLTs. Afterwards, it introduces Datachain, a lightweight, extensible and interoperable framework deliberately designed to ease the extraction of data hosted on DLTs. Through high-level query abstractions, users connect to underlying blockchains, perform data requests, extract transactions, manage data assets, and derive high-level analytic insights. Most importantly, due to the inherent interoperable nature of Datachain, queries and analytic jobs are reusable and can be executed without alterations on different underlying blockchains. To illustrate the wide applicability of Datachain, we present a realistic use-case on top of Hyperledger and BigchainDB.

The second paper is titled "Exploiting Twitter for Informativeness Classification in Disaster Situations" and authored by David Graf, Werner Retschitzegger, Wieland Schwinger, Birgit Pröll, and Elisabeth Kapsammer. It addresses the problem of disaster management. In essence, this matter urgently requires mechanisms for achieving situation awareness (SA) in a timely manner, allowing authorities to react in an appropriate way to reduce the impact on affected people and infrastructure. In such situations, no matter if they are human induced like shootings or natural ones like earthquakes or floods, social media platforms such as Twitter are frequently used communication channels, making them a highly valuable additional data source for enhancing SA. The challenge is, however, to identify out of the tremendous mass of irrelevant and non-informative social media data which messages are truly "informative", i.e., contributing to SA in a certain disaster situation. Existing approaches on machine learning driven informativeness classification most often focus on specific disaster types, such as shootings or floods, thus lacking general applicability and falling short in classification of new disaster events. Therefore, this paper puts forward the following three contributions: First, in order to better understand the underlying social media data source, an in-depth analysis of existing Twitter data on 26 different disaster events is provided along temporal, spatial, linguistic, and source dimensions. Second, based thereupon, a cross-domain informativeness classier is proposed, focussing not on specific disaster types but rather allowing for classifications across different types. Third, the applicability of this cross-domain classifier is demonstrated, showing its accuracy compared to other disaster type specific approaches.

In the third paper titled "COTILES: Leveraging Content and Structure for Evolutionary Community Detection," Nikolaos Sachpenderis, Georgia Koloniari, and Alexandros Karakasidis address community detection problems. Most related algorithms for online social networks rely solely either on the structure of the network or on its contents. Both extremes ignore valuable information that influences cluster formation. The authors propose COTILES, an evolutionary community detection algorithm, that leverages both structural and content-based criteria to derive densely connected communities with similar contents. Specifically, the authors extend a fast-online structural community detection algorithm by applying additional content-based constraints. They also further explore the effect of structure and content-based criteria on the clustering result by introducing three tunable variations of COTILES that either tighten or relax these criteria. Through an experimental evaluation, they show that the proposed method derives more cohesive communities compared to the original structured one and highlight when the proposed variations should be deployed.

Zahi Al Chami, Chady Abou Jaoude, Bechara Al Bouna, and Richard Chbeir propose "A Weighted Feature-Based Image Quality Assessment Framework in Real-Time" in the fourth paper. Nowadays, social media runs a significant portion of people's daily lives. Millions of people use social media applications to share photos. The massive volume of images shared on social media presents serious challenges and requires large computational infrastructure to ensure successful data processing. However, an image gets distorted somehow during the processing, transmission, sharing, or from a combination of many factors. So, there is a need to guarantee acceptable delivery content, especially for image processing applications. In this paper,

the authors present a framework developed to process a large number of images in real time while estimating the image quality. Image quality evaluation is measured based on four methods: Perceptual Coherence Measure, Semantic Coherence Measure, Content-Based Image Retrieval, and Structural Similarity Index. A weighted quality method is then calculated based on the four previous methods while providing a way to optimize the execution latency. Lastly, a set of experiments is conducted to evaluate the proposed approach.

In the fifth paper, "Sharing Knowledge in Digital Ecosystems Using Semantic Multimedia Big Data" is presented by Antonio M. Rinaldi and Cristiano Russo. In this paper, the authors stress the need to use formal representations in the context of multimedia big data due to the intrinsic complexity of this type of data. Furthermore, the relationships between objects should be clearly expressed and formalized to give the right meaning to the correlation of data. For this reason, the design of formal models to represent and manage information is a necessary task to implement intelligent information systems. Approaches based on the semantic web need to improve the data models that are the basis for implementing big data applications. Using these models, data and information visualization becomes an intrinsic and strategic task for the analysis and exploration of multimedia big data. In this paper, the authors propose the use of a semantic approach to formalize the structure of a multimedia big data model. Moreover, the identification of multimodal features to represent concepts and linguistic-semantic properties, relating them in an effective way, will bridge the gap between target semantic classes and low-level multimedia descriptors. The proposed model has been implemented in a NoSQL graph database populated by different knowledge sources. The authors explore a visualization strategy of this large knowledge base and present and discuss a case study for sharing information represented by a model according to a peer-to-peer (P2P) architecture. In this digital ecosystem, agents (e.g. machines, intelligent systems, robots, etc.) act like interconnected peers exchanging and delivering knowledge with each other.

"Facilitating and Managing Machine Learning and Data Analysis Tasks in Big Data Environments Using Web and Microservice Technologies" is proposed as the sixth paper by Shadi Shahoud, Sonja Gunnarsdottir, Hatem Khalloof, Clemens Duepmeier, and Veit Hagenmeyer. Here, the authors address the need for developing easy to use frameworks for instrumenting machine learning effectively for non-data analytics experts as well as novices. Furthermore, building machine learning models in the context of big data environments still represents a great challenge. In this paper, those challenges are addressed by introducing a new generic framework for efficiently facilitating the training, testing, managing, storing, and retrieving of machine learning models in the context of big data. The framework makes use of a powerful big data software stack platforms, web technologies, and a microservice architecture for a fully manageable and highly scalable solution. A highly configurable user interface hiding platform details from the user is introduced giving the user the ability to easily train, test, and manage machine learning models. Moreover, the framework automatically indexes and characterizes models and allows flexible exploration of them in the visual interface. The performance and usability of the new framework is evaluated on state-of-the-art machine learning algorithms: it is shown that executing, storing, and retrieving machine learning models via the framework results in an exceptionally low

overhead demonstrating that the framework can provide an efficient approach for facilitating machine learning in big data environments. Configuration options are also evaluated (e.g. caching of RDDs in Apache Spark) based on their affect runtime performance. Furthermore, the evaluation provides indicators for when the utilization of distributed computing (i.e., parallel computation) based on Apache Spark on a cluster outperforms single computer execution of a machine learning model.

The seventh paper is dedicated to "Stable Marriage Matching for Homogenizing Load Distribution in Cloud Data Center," authored by Disha Sangar, Ramesh Upreti, Harek Haugerud, Kyrre Begnum, and Anis Yazidi. Running a sheer virtualized data center with the help of Virtual Machines (VM) is the de facto standard in modern data centers. Live migration offers immense flexibility opportunities as it endows the system administrators with tools to seamlessly move VMs across physical machines. Several studies have shown that the resource utilization within a data center is not homogeneous across the physical servers. Load imbalance situations are observed when a significant portion of servers are either in overloaded or underloaded states. Apart from leading to an inefficient usage of energy by underloaded servers, this might lead to serious QoS degradation issues in the overloaded servers. In this paper, the authors propose a lightweight decentralized solution for homogenizing the load across different machines in a data center by mapping the problem to a Stable Marriage matching problem. The algorithm judiciously chooses pairs of overloaded and underloaded servers for matching and subsequently VM migrations are performed to reduce load imbalance. For the purpose of comparisons, three different greedy matching algorithms are also introduced. In order to verify the feasibility of the provided approach in real-life scenarios, the authors implement the solution on a small testbed. For the large-scale scenarios, they provide simulation results that demonstrate the efficiency of the algorithm and its ability to yield a near-optimal solution compared to other algorithms. The results are promising, given the low computational footprint of the algorithm.

The last paper of this special issue is titled "A Sentiment Analysis Software Framework for the Support of Business Information Architecture in the Tourist Sector" and is written by Javier Murga, Gianpierre Zapata, Heyul Chavez, Carlos Raymundo, Luis Rivera, Francisco Dominguez, Javier Moguerza, and José Maria Alvarez. It addresses a practical problem related to the increased use of digital tools within the Peruvian tourism industry, creating a corresponding increase in revenues. However, both factors have caused increased competition in the sector that in turn puts pressure on small and medium enterprises' (SME) revenues and profitability. This paper aims to apply neural network-based sentiment analysis on social networks to generate a new information search channel that provides a global understanding of user trends and preferences in the tourism sector. A working data-analysis framework is developed and integrated with tools from the cloud to allow a visual assessment of high probability outcomes based on historical data. This helps SMEs estimate the number of tourists arriving and places they want to visit, so that they can generate desirable travel packages in advance, reduce logistics costs, increase sales, and ultimately improve both quality and precision of customer service.

We hope this special issue motivates researchers to take the next step beyond building models to implement, evaluate, compare, and extend proposed approaches.

Many people worked long and hard to help this edition become a reality. We gratefully acknowledge and sincerely thank all the editorial board members and reviewers for their timely and valuable comments and insightful evaluations of the manuscripts that greatly improved the quality of the final versions. Of course, thanks go to all the authors for their contribution and cooperation. Finally, we thank the editors of TLDKS for their support and trust in us, and a special thanks to Gabriela Wagner for her availability and valuable work in the realization of this TLDKS volume.

July 2020

Richard Chbeir
Yannis Manolopoulos
Hiroshi Ishikawa
Sergio Ilarri
Apostolos Papadopoulos

Organization

Editors-in-Chief

Abdelkader Hameurlain	Paul Sabatier University, IRIT, France
A Min Tjoa	Technical University of Vienna, IFS, Austria

Guest Editors

Richard Chbeir	University of Pau and Adour Countries, France
Yannis Manolopoulos	Open University of Cyprus, Cyprus
Hiroshi Ishikawa	Tokyo Metropolitan University, Japan
Sergio Ilarri	University of Zaragoza, Spain
Apostolos Papadopoulos	Aristotle University of Thessaloniki, Greece

Editorial Board

Reza Akbarinia	Inria, France
Dagmar Auer	FAW, Austria
Djamal Benslimane	University Lyon 1, France
Stéphane Bressan	National University of Singapore, Singapore
Mirel Cosulschi	University of Craiova, Romania
Dirk Draheim	University of Innsbruck, Austria
Johann Eder	Alpen Adria University of Klagenfurt, Austria
Anastasios Gounaris	Aristotle University of Thessaloniki, Greece
Theo Härder	Technical University of Kaiserslautern, Germany
Sergio Ilarri	University of Zaragoza, Spain
Petar Jovanovic	Universitat Politècnica de Catalunya, BarcelonaTech, Spain
Aida Kamišalić Latifić	University of Maribor, Slovenia
Dieter Kranzlmüller	Ludwig-Maximilians-Universität München, Germany
Philippe Lamarre	INSA Lyon, France
Lenka Lhotská	Technical University of Prague, Czech Republic
Vladimir Marik	Technical University of Prague, Czech Republic
Jorge Martinez Gil	Software Competence Center Hagenberg, Austria
Franck Morvan	Paul Sabatier University, IRIT, France
Torben Bach Pedersen	Aalborg University, Denmark
Günther Pernul	University of Regensburg, Germany
Soror Sahri	LIPADE, Descartes Paris University, France
Shaoyi Yin	Paul Sabatier University, France
Feng “George” Yu	Youngstown State University, USA

SI Editorial Board

Sabri Allani	Université de Pau et des Pays de l'Adour, France
Adel Alti	Constantine University, Algeria
Richard Chbeir	Université de Pau et des Pays de l'Adour, France
Joyce El Haddad	UDO, France
Anna Formica	Istituto di Analisi dei Sistemi ed Informatica, CNR, Italy
Anastasios Gounaris	Aristotle University of Thessaloniki, Greece
Michael Granitzer	University of Passau, Germany
Abdelkader Hameurlain	Paul Sabatier University, France
Ramzi Haraty	Lebanese American University, Lebanon
Masaharu Hirota	Okayama University of Science, Japan
Sergio Ilarri	University of Zaragoza, Spain
Hiroshi Ishikawa	Tokyo Metropolitan University, Japan
Lara Kallab	Nobatek, France
Helen Karatza	Aristotle University of Thessaloniki, Greece
Georgia Koloniari	University of Macedonia, Greece
Anne Laurent	University of Montpellier, France
Aristidis Likas	University of Ioannina, Greece
Yannis Manolopoulos	Open University of Cyprus, Cyprus
Apostolos Papadopoulos	Aristotle University of Thessaloniki, Greece
Imad Saleh	University of Paris 8, France
Maria Luisa Sapino	Università degli Studi di Torino, Italy
Joe Tekli	Lebanese American University, Lebanon
Demetris Trihinas	University of Nicosia, Cyprus
Jose R. R. Viqueira	University of Santiago de Compostela, Spain

External Reviewers

Karam Bou Chaaya	Université de Pau et des Pays de l'Adour, France
Elio Mansour	Université de Pau et des Pays de l'Adour, France

Contents