

Big Data Analytics for Time-Critical Mobility Forecasting

George A. Vouros • Gennady Andrienko •
Christos Doulkeridis • Nikolaos Pelekis •
Alexander Artikis • Anne-Laure Joussetme •
Cyril Ray • Jose Manuel Cordero • David Scarlatti
Editors

Big Data Analytics for Time-Critical Mobility Forecasting

From Raw Data to Trajectory-Oriented
Mobility Analytics in the Aviation and
Maritime Domains

 Springer

Editors

George A. Vouros
Department of Digital Systems
University of Piraeus
Piraeus, Greece

Gennady Andrienko
Intelligent Analysis & Info Systems
Fraunhofer Institute IAIS
Sankt Augustin, Germany

Christos Doukeridis
Department of Digital Systems
University of Piraeus
Piraeus, Greece

Nikolaos Pelekis
Department of Statistics and Insurance
Science
University of Piraeus
Piraeus, Greece

Alexander Artikis
Complex Event Recognition Group
NCSR “Demokritos”
Agia Paraskevi, Greece

Anne-Laure Jousselme
Centre for Maritime Research and
Experimentation (CMRE)
NATO Science and Technology
Organization (STO)
La Spezia, Italy

Cyril Ray
Naval Research Institute (IRENav)
Arts & Metiers-Paris Tech
Brest, France

Jose Manuel Cordero
Edificio Allende
CRIDA
Madrid, Spain

David Scarlatti
Boeing Research & Technology Europe
Madrid, Spain

ISBN 978-3-030-45163-9 ISBN 978-3-030-45164-6 (eBook)
<https://doi.org/10.1007/978-3-030-45164-6>

© Springer Nature Switzerland AG 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 Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

*This book is dedicated to all those who
struggle for a better own and others'
trajectory with philotimo,¹ respect and with
no obsession.*

¹Philotimo (or filotimo) is a Greek word, which is difficult to translate. One may start from here, among other references: <https://en.wikipedia.org/wiki/Philotimo>.

Preface

Spatiotemporal mobility data has a significant role and impact on the global economy and our everyday lives. The improvements along the last decades in terms of data management, planning of operations, security of operations, information provision to operators and end-users have been driven by location-centered information. While a shift of paradigm regarding mobility data towards trajectory-oriented tasks is emerging in several domains, the ever-increasing volume of data emphasizes the need for advanced methods supporting detection and prediction of events and trajectories, supplemented by advanced visual analytic methods, over multiple heterogeneous, voluminous, fluctuating, and noisy data streams of moving entities. This book provides a comprehensive and detailed description of Big Data solutions towards activity detection and forecasting in very large numbers of moving entities spread across large geographical areas. Specifically, following a trajectory-oriented approach, this book reports on the state-of-the-art methods for the detection and prediction of trajectories and important events related to moving entities, together with advanced visual analytics methods, over multiple heterogeneous, voluminous, fluctuating, and noisy data streams from moving entities, correlating them with data from archived data sources expressing, among others, entities' characteristics, geographical information, mobility patterns, regulations, and intentional data (e.g., planned routes), in a timely manner. Solutions provided are motivated, validated, and evaluated in user-defined challenges focusing on increasing the safety, efficiency, and economy of operations concerning moving entities in the air-traffic management and maritime domains.

The book contents have been structured into six parts:

The first part provides the motivating points and background for mobility forecasting supported by trajectory-oriented analytics. It presents specific problems and challenges in the aviation (air-traffic management) and the maritime domains and clarifies operational concerns and objectives in both domains. It presents domain-specific terminology used in examples and cases, in which technology is demonstrated, evaluated/validated, throughout the book. Equally important to the above is the presentation of the data sources exploited per domain, the big data challenges ahead in both domains, and of course, the requirements from

technologies presented in subsequent parts of the book. These chapters present data exploited for operational purposes in the aviation and maritime domains and provide an initial understanding of spatiotemporal data through specific examples. They also present challenges and motivating points by means of operational scenarios where technology can help, putting the technologies presented in subsequent parts of the book in a unique frame: This helps us understand why technological achievements are necessary, what are the domain-specific requirements driving developments in analytics, data storage, and processing, and what are the data processing, data management, and data-driven analytics tools needed to advance operational goals towards trajectory-based operations.

The second part focuses on big data quality assessment and processing, as applied in the data sources and according to the requirements and objectives presented in the first part of the book. This, second part of the book, presents novel technologies, appropriate to serve mobility analytics components that are presented in subsequent sections. In doing so, workflows regarding data sources' quality assessment via visual analytics methods are considered to be essential to understand inherent features and imperfections of data, affecting the ways data should be processed and managed, as the first section of this part shows. In addition to this, methods for online construction of streamed data synopses are presented, towards addressing big data challenges presented by surveillance, mostly, data sources.

The third part of this book specifies solutions towards managing big spatiotemporal data: The first section specifies a generic ontology revolving around the notion of trajectory so as to model data and information that is necessary for analytics components. This ontology provides a generic model for constructing knowledge graphs integrating data from disparate data sources. In conjunction to this, this part describes novel methods for integrating data from archived and streamed data sources. Special emphasis is given to enriching data streams and integrating streamed and archival data to provide coherent views of mobility: This is addressed by real-time methods discovering topological and proximity relations among spatiotemporal entities. Finally, distributed storage of integrated dynamic and archived mobility data—i.e. large knowledge graphs constructed according to the generic model introduced—are within focus.

The next part focuses on mobility analytics methods exploiting (online) processed, synopsized, and enriched data streams as well as (offline) integrated, archived mobility data. Specifically, online future location prediction methods and trajectory prediction methods are presented, distinguishing between short-term and the challenging long-term predictions. Recognition of complex events in challenging cases for detecting complex events is thoroughly presented. In addition to this, an industry-strong maritime anomaly detection service capable of processing daily real-world data volumes is presented. This part focuses also on offline trajectory analytics, addressing trajectory clustering and detection of routes followed by mobile entities. Novel algorithms for subtrajectory clustering are proposed and evaluated.

The fifth part presents how methods addressing data management, data processing, and mobility analytics are integrated in a big data architecture that

has distinctive characteristics when compared to known big data paradigmatic architectures. We call this architectural paradigm, which is based on well-defined principles for building analytics pipelines δ . This paradigm is instantiated to a specific architecture realizing the datAcron integrated system prototype. This part presents the software stack of the datAcron system, together with issues concerning individual, online, and offline components integration.

The last part focuses on important ethical issues that research on mobility analytics should address: This is deemed to be crucial, given the growth of interest in that topic in computer science and operational stakeholders, necessitating the sharing of data and distributing the processing among stakeholders.

All chapters present background information on the specific topics they address, detailed and rigorous specification of scientific and technological problems considered, and state-of-the-art methods addressing these problems, together with novel approaches that authors have developed, evaluated, and validated, mainly during the last 3 years of their involvement in the datAcron H2020 ICT Big Data Project. Evaluation and validation results per method are presented using data sets from both, maritime and aviation domain, showing the potential and the limitations of methods presented, also according to the requirements specified in the first part of the book. The chapters present also technical details about implementations of methods, aiming to address big data challenges, so as to achieve the latency and throughput requirements set in both domains.

In doing so, this book aims to present a reference book to all stakeholders in different domains with mobility detection and forecasting needs and computer science disciplines aiming to address data-driven mobility data exploration, processing, storage, and analysis problems.

I would like to take the opportunity to thank everybody who contributed to the exciting effort of developing mobility data processing, storage, analysis solutions in time-critical domains, whose state of the art is summarized in this book. These, as part of a much wider community, include all co-editors and chapter authors of this publication. This book is a concerted effort of many people who worked and continue to work together in different, but always exciting, lines of research for mobility analytics.

Piraeus, Greece
February 2020

George A. Vouros

Acknowledgements

The developments described in the chapters of this book have been developed in the course of work in several past and on-going research projects, whose support is also acknowledged by the authors of each chapter. However, ideas and large part of this work have been developed, evaluated, and validated, mainly during the 3 years of involvement in the datAcron H2020 ICT Big Data Project.

datAcron has been funded by the European Union's Horizon 2020 Programme under grant agreement No. 687591. datAcron is a research and innovation collaborative project whose aim was to introduce novel methods to detect threats and abnormal activity of very large numbers of moving entities in large geographic areas.

Towards this target, datAcron advanced the management and integrated exploitation of voluminous and heterogeneous data-at-rest (archival data) and data-in-motion (streaming data) sources, so as to significantly advance the capacities of systems to promote safety and effectiveness of critical operations for large numbers of moving entities in large geographical areas.

Technological developments in datAcron have been validated and evaluated in user-defined challenges that aim at increasing the safety, efficiency, and economy of operations concerning moving entities in the air-traffic management (ATM) and maritime domains.

The datAcron addressed the following core challenges:

- **Distributed management and querying of integrated spatiotemporal RDF data-at-rest and data-in-motion in integrated manners:** datAcron advanced RDF data processing and spatiotemporal query answering for very large numbers of real-world triples and spatiotemporal queries, providing also native support for trajectory data, handling (semantic) trajectories as first-class citizens in data processing. In situ data processing and link discovery for data integration are critical technologies to those targets.
- **Detection and prediction of trajectories of moving entities in the aviation and maritime domains:** datAcron developed novel methods for real-time trajectory reconstruction, aiming at efficient large-scale mobility data analytics. Real-time

trajectories forecasting for the aviation and maritime domains aim to a short forecasting horizon.

- **Recognition and forecasting of complex events in the aviation and maritime domains:** datAcron developed methods for event recognition under uncertainty in noisy settings, aiming at processing very large number of events/second with complex event definitions. In doing so, optimization of complex events patterns' structure and parameters by means of machine learning methods for constructing event patterns was within datAcron objectives.
- **Visual analytics in the aviation and maritime domains:** datAcron developed a general visual analytics infrastructure supporting all steps of analysis through appropriate interactive visualizations, including both generic components and components tailored for specific applications.

Contents

**Part I Time Critical Mobility Operations and Data: A Perspective
from the Maritime and Aviation Domains**

1	Mobility Data: A Perspective from the Maritime Domain	3
	Cyril Ray, Anne-Laure Joussetme, Clément Iphar, Maximilian Zocholl, Elena Camossi, and Richard Dréo	
1.1	Maritime Operational Scenarios: Challenges and Requirements	3
1.1.1	Monitoring Fishing Activities.....	4
1.1.1.1	Secured Fishing	5
1.1.1.2	Maritime Sustainable Development.....	8
1.1.1.3	Maritime Security.....	9
1.1.2	Maritime Situational Indicators	11
1.2	Big Mobility Data in the Maritime Domain	11
1.2.1	Maritime Big Data Challenges	13
1.2.1.1	Variety	13
1.2.1.2	Veracity	14
1.2.1.3	Volume	14
1.2.1.4	Velocity	14
1.2.1.5	Context	15
1.2.2	Heterogeneous Integrated Dataset for Maritime ISR	16
1.2.2.1	Navigation-Related Data	18
1.2.2.2	Vessel Data	22
1.2.2.3	Geographic Data	23
1.2.2.4	Environmental Data.....	26
1.2.3	Generating Operational Scenario for Experiments.....	28
1.3	Conclusions.....	28
	References.....	30
2	The Perspective on Mobility Data from the Aviation Domain	33
	Jose Manuel Cordero and David Scarlatti	
2.1	Introduction.....	33

2.2	Trajectory Prediction Approaches in the Aviation Domain	35
2.2.1	Kinematic Trajectory Prediction Approach	35
2.2.2	Kinetic Trajectory Prediction Approach	36
2.2.3	Data-Driven Trajectory Prediction Approaches	37
2.3	Aviation Datasets	41
2.4	Reconstructed Trajectory	44
2.4.1	Aircraft Intent	44
2.4.2	The Trajectory Reconstruction Process	48
2.5	Aviation Operational Scenarios: Big Data Challenges and Requirements	49
2.5.1	Regulations Detection and Prediction	52
2.5.2	Demand and Capacity Imbalance Detection and Prediction	53
2.5.3	Trajectory Prediction: Preflight	53
2.5.4	Trajectory Prediction: Real Time	54
2.6	Conclusions	54
	References	55

Part II Visual Analytics and Trajectory Detection and Summarization: Exploring Data and Constructing Trajectories

3	Visual Analytics in the Aviation and Maritime Domains	59
	Gennady Andrienko, Natalia Andrienko, Georg Fuchs, Stefan Rüping, Jose Manuel Cordero, David Scarlatti, George A. Vouros, Ricardo Herranz, and Rodrigo Marcos	
3.1	Introduction	60
3.2	Related Work	60
3.3	Visual Exploration of Data Quality	62
3.4	Examples of Visual Analytics Processes	67
3.4.1	Detection and Analysis of Anchoring Events in Maritime Traffic	68
3.4.2	Exploring Separation of Airport Approach Routes	72
3.4.3	Revealing Route Choice Criteria of Flight Operators	75
3.4.4	Understanding Airspace Configuration Choices	77
3.5	Discussion and Conclusion	81
	References	82
4	Trajectory Detection and Summarization over Surveillance Data Streams	85
	Kostas Patroumpas, Eva Chondrodima, Nikos Pelekis, and Yannis Theodoridis	
4.1	Introduction	86
4.2	Related Work	89
4.3	Streaming Data Sources in Maritime and Aviation Surveillance	91

4.3.1	Maritime Data Sources	91
4.3.2	Aviation Data Sources	92
4.4	System Overview	93
4.4.1	Trajectory Representation	94
4.4.2	Framework Architecture	95
4.5	Online Processing of Streaming Trajectories	96
4.5.1	Trajectory Construction	97
4.5.1.1	Noise Reduction	97
4.5.1.2	Mobility State Maintenance	99
4.5.2	Trajectory Summarization.....	100
4.5.2.1	Mobility Events on (x, y) Dimensions	101
4.5.2.2	Mobility Events on z -Dimension.....	103
4.5.2.3	Mobility Events on t -Dimension.....	105
4.5.2.4	Discussion.....	105
4.6	Prototype Implementation	106
4.6.1	Custom Support for Mobility Features	107
4.6.2	Processing Through a Pipeline of Operators	108
4.7	Empirical Validation	109
4.7.1	Performance Results.....	110
4.7.2	Approximation Error	112
4.7.3	Comparison with Trajectory Simplification Methods.....	113
4.8	Towards Cross-Stream Trajectory Maintenance	115
4.8.1	Methodology.....	115
4.8.2	Proof-of-Concept Evaluation	116
4.9	Conclusions.....	117
	References.....	118

Part III Trajectory Oriented Data Management for Mobility Analytics

5	Modeling Mobility Data and Constructing Large Knowledge Graphs to Support Analytics: The datAcron Ontology	123
	Georgios M. Santipantakis, George A. Vouros, Akrivi Vlachou, and Christos Doulkeridis	
5.1	Introduction.....	123
5.2	Requirements for Enriched Representation of Mobility Data.....	125
5.2.1	Requirements for the Representation of Semantic Trajectories.....	125
5.2.2	Fundamental Data Types and Data Transformations for Visual Analytics	128
5.3	Semantic Representations of Trajectories: Related Work	130
5.4	The datAcron Ontology	132
5.4.1	Core Vocabulary and Overall Structure.....	133
5.4.2	Patterns for the Representation of Semantic Trajectories.....	134

5.5	Supporting Data Transformations	140
5.5.1	Data Sets	140
5.5.2	datAcron Namespaces for Functions	142
5.5.3	Validation Setup	143
5.5.4	Pre-processing Steps and Auxiliary Structures.....	144
5.5.5	Visual Analytics Enhanced Via Data Transformations on Aviation Use Cases.....	145
5.5.6	Visual Analytics Enhanced Via Data Transformations on Maritime Use Cases	148
5.6	Conclusions.....	150
	References.....	151
6	Integrating Data by Discovering Topological and Proximity Relations Among Spatiotemporal Entities	155
	Georgios M. Santipantakis, Christos Doulkeridis, Akrivi Vlachou, and George A. Vouros	
6.1	The Role of Link Discovery in Data Integration.....	155
6.2	Background on Spatiotemporal Link Discovery	156
6.2.1	Principles of Link Discovery.....	157
6.2.2	The Problem of Spatiotemporal Link Discovery	157
6.2.2.1	Topological Relations.....	158
6.2.2.2	Proximity Relations.....	158
6.2.2.3	Temporal Relations	158
6.2.2.4	Motivating Real-Life Applications and Examples	159
6.2.3	Existing Spatial Link Discovery Systems and Frameworks	160
6.3	A New Spatiotemporal LD Framework	161
6.3.1	The Architecture of a Spatiotemporal Link Discovery Framework	161
6.3.1.1	Connectors and Configurations	163
6.3.1.2	Consumption Strategy	164
6.3.1.3	Blocking Methods (Filter).....	165
6.3.1.4	Link Evaluation	165
6.3.2	Link Discovery of Topological Relations	166
6.3.2.1	A Baseline Link Discovery Algorithm.....	166
6.3.2.2	The RADON Algorithm	167
6.3.3	The MaskLink Technique for Link Discovery of Topological Relations	168
6.3.4	Link Discovery of Proximity Relations.....	170
6.3.5	Refined Blocking Method	171
6.4	Scalable LD for Spatiotemporal Data	172
6.4.1	Setup for Scalable Link Discovery	172
6.4.2	Stream-Based Link Discovery	173

6.5	Empirical Evaluation	175
6.5.1	Experimental Setup	175
6.5.2	Results for Region-to-Region Topological Relations	176
6.5.3	Results for Point-to-Region Proximity-Based Relations	177
6.6	Conclusions	178
	References	179
7	Distributed Storage of Large Knowledge Graphs with Mobility Data	181
	Panagiotis Nikitopoulos, Nikolaos Koutroumanis, Akrivi Vlachou, Christos Doukeridis, and George A. Vouros	
7.1	Introduction	181
7.2	Background	182
7.2.1	Distributed Processing Frameworks	183
7.2.2	Scalable Querying of Large Knowledge Graphs	184
7.3	One-Dimensional Encoding	185
7.4	Overview of the DiStRDF Engine	187
7.5	The Storage Layer	188
7.5.1	Storing the Dictionary	188
7.5.2	Storing RDF Triples	189
7.6	The Processing Layer	191
7.6.1	The SPARQL Query Parsing Component	191
7.6.2	The Logical Plan Builder	191
7.6.3	The Logical Plan Optimizer	194
7.6.4	The Physical Plan Constructor	196
7.6.4.1	Datasource Operators	196
7.6.4.2	Selection Operators	197
7.6.4.3	Join Operators	198
7.6.4.4	Projection Operator	199
7.6.4.5	Spatiotemporal Operators	199
7.6.4.6	Other Operators	200
7.7	Experimental Evaluation	200
7.7.1	Experimental Setup	200
7.7.2	Experiments on Real-World Queries	201
7.7.2.1	Queries for Maritime and Aviation Domain ...	202
7.7.2.2	Evaluation	204
7.7.3	Experiments on Physical Query Optimization	205
7.7.4	Experiments on Logical Query Optimization	207
7.7.5	Experiments on Spatiotemporal Filtering	207
7.8	Conclusions	209
	References	210

Part IV Analytics Towards Time Critical Mobility Forecasting

8	Future Location and Trajectory Prediction	215
	Harris Georgiou, Petros Petrou, Panagiotis Tampakis, Stylianos Sideridis, Eva Chondrodima, Nikos Pelekis, and Yannis Theodoridis	
8.1	Introduction	215
8.1.1	Purpose, Scope, Motivation	215
8.2	Background	217
8.2.1	Future Location Prediction	217
8.2.2	Trajectory Prediction	219
8.2.3	Related Datasets Used	220
8.3	Distributed Online Future Location Prediction (FLP)	221
8.3.1	Short-Term FLP: Routes-Agnostic Approach	223
8.3.1.1	Long Short-Term Memory (LSTM)	223
8.3.2	Long-Term FLP: Routes-Based Approach	226
8.3.2.1	Offline Step: Routes Network Discovery (Medoids)	226
8.3.2.2	Online Step: Network-Based Prediction	228
8.3.3	Experimental Study	229
8.3.3.1	Short-Term FLP: Routes-Agnostic Approach	229
8.3.3.2	Long-Term FLP: Routes-Based Approach	231
8.4	Semantics-Driven Trajectory Prediction (FSTP)	232
8.4.1	Clustering Enriched Trajectories for FSTP	234
8.4.2	Predictive Models	236
8.4.2.1	Hidden Markov Models (HMM)	238
8.4.2.2	Linear Regressors (LR)	240
8.4.2.3	Regression Trees (CART)	241
8.4.2.4	Feed-Forward Neural Network Regressors (NN-MLP)	241
8.4.3	Experimental Study	242
8.4.3.1	Clustering Stage	242
8.4.3.2	Predictive Modeling Stage	243
8.5	Related Work	247
8.6	Discussion	250
8.7	Conclusion	251
	References	252
9	Event Processing for Maritime Situational Awareness	255
	Manolis Pitsikalis, Konstantina Bereta, Marios Vondas, Dimitris Zissis, and Alexander Artikis	
9.1	Introduction	255
9.2	System Architecture	257
9.2.1	Setting the Scene	257
9.2.2	Maritime Stream Reasoning	258

9.3	Maritime Events	260
9.3.1	Building Blocks	261
9.3.1.1	Vessel Within Area of Interest	261
9.3.1.2	Communication Gap	262
9.3.2	Maritime Situational Indicators	263
9.3.2.1	Vessel with High Speed Near Coast	263
9.3.2.2	Anchored or Moored Vessel	264
9.3.2.3	Drifting Vessel	266
9.3.2.4	Tugging	267
9.3.2.5	Vessel rendez-vous	267
9.4	Anomaly Detection Service	269
9.5	Summary & Further Work	273
	References	273
10	Offline Trajectory Analytics	275
	Panagiotis Tampakis, Stylianos Sideridis, Panagiotis Nikitopoulos, Nikos Pelekis, Christos Doukeridis, and Yannis Theodoridis	
10.1	Introduction	275
10.1.1	About Mobility Data Analytics	276
10.2	Background	277
10.2.1	(Sub)trajectory Clustering	277
10.2.2	Hotspot Analysis	278
10.2.3	Data-Enriched Mobility Networks	278
10.3	Distributed Trajectory Clustering	279
10.3.1	Distributed Whole-Trajectory Clustering	279
10.3.1.1	Clustering Trajectories in a Distributed Way with Spark MLlib	279
10.3.2	Distributed Subtrajectory Clustering	281
10.3.2.1	Definitions	282
10.3.2.2	Distributed Subtrajectory Join	285
10.3.2.3	Distributed Trajectory Segmentation	285
10.3.2.4	Distributed Subtrajectory Clustering	288
10.3.2.5	Experimental Results	289
10.4	Distributed Hotspot Analysis	292
10.4.1	Definitions	292
10.4.2	Exact THS Algorithm	294
10.4.3	An Approximate Algorithm: aTHS	294
10.4.4	Experimental Study	296
10.4.4.1	Experimental Results	297
10.5	Distributed Data-Enriched Mobility Networks	298
10.5.1	Definitions	298
10.5.2	Discovering Data-Enriched Mobility Networks	299
10.5.2.1	Step 1: Enriched Nodes Extraction	300
10.5.2.2	Step 2: Enriched Paths Discovery	300
10.5.3	The SeaAirNet Algorithm	302

10.5.4	Experimental Results	303
10.5.4.1	Datasets	304
10.5.4.2	Qualitative Results in Aviation	304
10.5.4.3	Qualitative Results in Maritime	305
10.6	Related Work	306
10.7	Discussion: Lessons Learnt	310
	References	311

Part V Big Data Architectures for Time Critical Mobility Forecasting

11	The δ Big Data Architecture for Mobility Analytics	315
	George A. Vouros, Apostolis Glenis, and Christos Doukeridis	
11.1	Introduction	315
11.2	Background, Motivating Points, and Requirements	317
11.2.1	Motivating Points and Background	317
11.2.2	Requirements	318
11.2.3	Related Work	320
11.3	The δ Big Data Architecture	321
11.3.1	The δ Architecture: An Overview	321
11.3.2	Principles and Rules	322
11.3.3	Snapshots of the δ Architecture	327
11.3.3.1	A Single Log, No Data Management Configuration	327
11.3.3.2	Multiple Logs, No Data Management Configuration	328
11.3.3.3	The λ Architecture in Multiple Layers	328
11.3.3.4	The κ Architecture in Multiple Layers Configuration	328
11.4	Instantiating δ in the datAcron Big Data System for Mobility Analytics	329
11.5	Implementation of datAcron Integrated Prototype	332
11.5.1	Information Processing and Data Management	332
11.5.2	Online Future Location Prediction and Trajectory Prediction	333
11.5.3	Complex Event Recognition/Forecasting Online	333
11.5.4	Visualization and Interactive Visual Analytics: Online ...	334
11.5.5	Trajectory Data Analytics Offline	334
11.5.6	Complex Event Recognition/Forecasting Offline	335
11.5.7	Interactive Visual Analytics Offline	335
11.6	Experimental Evaluation	336
11.7	Concluding Remarks	338
	References	338

Part VI Ethical Issues for Time Critical Mobility Analytics

12 Ethical Issues in Big Data Analytics for Time Critical Mobility Forecasting	343
Gemma Galdon Clavell and Victoria Peuvrelle	
12.1 Introduction	343
12.2 Ethics in Research at the EU Level	344
12.3 Ethical Issues	346
12.3.1 Consequences of Outputs for Privacy	346
12.3.2 Function Creep and Dual Use	347
12.3.3 Other Ethics Issues	348
12.3.3.1 Subjective Thresholds	348
12.3.3.2 Algorithms and Automatic Decision-Making	348
12.3.3.3 Accuracy of Data	349
12.3.3.4 Undesirable Reuse and Threat Modeling	349
12.3.3.5 Technological Divide and Discrimination	349
12.4 Tackling Ethics Issues	350
12.4.1 Identification of Factors for Privacy Concerns	350
12.4.1.1 Types of Moving Entities Involved	350
12.4.1.2 Provenance and Type of Data	351
12.4.1.3 Actors/End-Users	351
12.4.2 Identification of Factors to Avoid Function Creep	352
12.4.3 The Integration of Privacy by Design Within the Process	352
12.4.4 Drafting a Data Management Plan	355
12.4.5 Identification and Documentation of the Licensing Options	355
12.5 Conclusion	357
References	358
Index	359

Editors and Contributors

About the Editors

Dr. Gennady Andrienko is a lead scientist responsible for the visual analytics research at Fraunhofer Institute IAIS (Sankt Augustin, Germany) and full professor at City University London. He co-authored monographs “Exploratory Analysis of Spatial and Temporal Data” (Springer, 2006) and “Visual Analytics of Movement” (Springer, 2013) and about 100 peer-reviewed journal papers. He is associate editor of three journals: “Information Visualization” (since 2012), “IEEE Transactions on Visualization and Computer Graphics” (2012–2016), and “International Journal of Cartography” (since 2014) and editorial board member of “Cartography and Geographic Information Science” and “Cartographica.” He received Test of Time award at IEEE VAST 2018 and best paper awards at AGILE 2006, IEEE VAST 2011 and 2012 conferences, and EuroVA 2018 workshop.

Alexander Artikis is an Assistant Professor in the Department of Maritime Studies of the University of Piraeus. He is also a Research Associate in the Institute of Informatics and Telecommunications of NCSR Demokritos, the largest research center in Greece, leading the Complex Event Recognition lab. His research interests lie in the field of Artificial Intelligence. He has published over 100 papers in the top conferences and journals of the field, while, according to Google Scholar, his h-index is 31. He has participated in several EU-funded Big Data projects, being the scientific coordinator in some of them. He has been serving a member of the program committee of various conferences, such as AAAI, IJCAI, AAMAS, ECAI, KR, and DEBS.

Jose Manuel Cordero holds the Telecommunications Engineer degree from the Universidad de Sevilla, Spain (2002). He is currently a Principal Researcher at CRIDA (ATM R&D Reference Centre, depending on the Spanish ANSP, ENAIRE), with over 15 years of experience in the air-traffic management domain in the areas of performance monitoring and assessment, system simulation, and validation. In the

last years, he focused his activity on Performance Management projects, including research activities in big data analytics, predictive models, and multi-objective optimization methods.

Christos Doukeridis is an Assistant Professor at the Department of Digital Systems in the University of Piraeus. He has been awarded both a Marie-Curie fellowship and an ERCIM “Allain Bensoussan” fellowship for postdoctoral studies at the Norwegian University of Science and Technology in 2011 and 2009, respectively. He was the Principal Investigator of the research project “RoadRunner: Scalable and Efficient Analytics for Big Data” (2014–2015, funded by the General Secretariat for Research and Technology in Greece). He has participated in several H2020 EU research projects related to Big Data management and analytics (“Track&Know” 2018–2020, “BigDataStack,” 2018–2020, and “datAcron,” 2016–2018). He has been awarded the first position in the 2017 SemEval challenge on Sentiment Analysis in Twitter and the third position in the 2016 ACM SIGSPATIAL Cup on Hot Spot Analysis of Mobility Data. He has published in top international journals (including VLDB Journal, IEEE TKDE, IEEE JSAC, ACM TKDD, Data Mining and Knowledge Discovery, Information Systems, Distributed and Parallel Databases) and conferences (including ACM SIGMOD, VLDB, ICDE, EDBT, SSTD, PKDD, SIAM SDM) in the areas of data management, knowledge discovery, and distributed systems.

Dr. Anne-Laure Jousselme is with the NATO Centre for Maritime Research and Experimentation. She is member of the Boards of Director of the International Society of Information Fusion and Belief Functions and Applications Society. She is associate editor of the Perspectives on Information Fusion magazine and area editor of the International Journal of Approximate Reasoning. Her research interests include maritime anomaly detection, information fusion, reasoning under uncertainty, information quality.

Nikos Pelekis is Assistant Professor at the Department of Statistics and Insurance Science, University of Piraeus, Greece. His research interests include all topics of data science. He has been particularly working for almost 20 years in the field of Mobility Data Management and Mining. Nikos has co-authored one monograph and more than 80 refereed articles in scientific journals and conferences, receiving more than 1000 citations, while he has received 3 best paper awards and won the SemEval’17 competition and ranked 3rd in ACM SIGSPATIAL’16 data challenge. He has offered several invited lectures in Greece and abroad (including PhD/MSc/summer courses at Rhodes, Milano, KAUST, Aalborg, Trento, Ghent, JRC Ispra) on Mobility Data Management and Data Mining topics. He has been actively involved in more than 10 European and National R&D projects. Among them, he is or was principal researcher in GeoPKDD, MODAP, MOVE, DATASIM,

SEEK, DART and datAcron, Track & Know, MASTER. For more information: <http://www.unipi.gr/faculty/npelekis/>.

Cyril Ray is associate professor in computer science at Arts & Metiers—ParisTech, attached to Naval Academy Research Institute (IRENav) in France. His current research is oriented to the modeling and design of location-based services. His work mainly concerns theoretical aspects of the design of ubiquitous and adaptive location-based services applied to human mobility, maritime, and urban transportation systems. This research addresses the relationship between geographic information systems and the underlying computing architectures that support real-time tracking of mobile objects (pedestrian in indoor spaces, vehicles in urban areas, and ships at sea). This work includes, at different level, integration of location acquisition technologies, modeling of heterogeneous and large spatiotemporal datasets, movement data processing (cleaning, filtering, trajectory modeling, knowledge discovery), modeling of context-aware systems, and traffic simulation and prediction.

David Scarlatti works as Data Solutions Architect at Boeing Research & Technology Europe in the Aerospace Operational Efficiency group. He received Aeronautics Engineering degree in 1994 (Universidad Politecnica de Madrid), is Stanford Certified Project Manager (2008), Master of Science in Technology Management (2010, Open University- UK), and GIAC Certified Incident Handler (2011). In 1989, he started to work with computers at PHILIPS, then worked at INDRA (1993–2000) as Software Engineer, and at ORACLE as Technology Manager (2000–2005) where he led the European Professional Community on Data Warehousing. In May 2005, he joined Boeing where he has been applying data analytics technologies to a wide variety of aviation-related problems. His fields of expertise include big data, visualization, advanced computing in data analysis, cybersecurity, and human-machine interfaces. Is co-inventor in 17 patents (5 granted).

George A. Vouros holds a BSc in Mathematics (1986) and a PhD in Artificial Intelligence (1992) all from the University of Athens, Greece. Currently, he is a Professor in the Department of Digital Systems in the University of Piraeus and head of the AI-Lab in this Department. He has done research in the areas of Expert Systems, Knowledge Management, Collaborative Systems, Ontologies, and Agents and Multi-Agent Systems. He served/serves as program chair, chair, and member of organizing committees of national and international conferences (AAMAS, AAI, IJCAI, ECAI, WI/IAT, AT, EUMAS, ICMLA, ESWC, CSCL, AIAI, ISWC) and as member of steering committees/boards of international conferences/workshops (EURAMAS, SETN, AT, COIN, OAEI). He has given keynote speeches in conferences and workshops (WoMo, ICTAI, CLIMA, IF&GIS) and he has organized several workshops (MATES@EDBT, Data-Driven ATM@WAC, Data-Enhanced Trajectory-Based Operations@ICRAT; the most recent ones). He served/ serves as guest editor in special issues in well-reputed journals (e.g., IJCIS,

AIR, GEOINFORMATICA, AICom, ISF). He is/was senior researcher in numerous EU-funded and National research projects (GSRT/ AMINESS, FP7/Grid4All, FP7/SEMAGROW, COST/Agreement Technologies the most recent ones). He recently coordinated the successful DART (SESARER) project on data-driven trajectory prediction in the aviation domain and coordinates the datAcron (H2020 ICT-16) Big Data project. For more information on recent activities and publications, see <http://ai-group.ds.unipi.gr/georgev/>.

Contributors

Gennady Andrienko Fraunhofer Institute IAIS, Sankt Augustin, Germany
City University of London, London, UK

Natalia Andrienko Fraunhofer Institute IAIS, Sankt Augustin, Germany
City University of London, London, UK

Alexander Artikis Department of Maritime Studies, University of Piraeus, Piraeus, Greece
Institute of Informatics & Telecommunications, NCSR Demokritos, Athens, Greece

Konstantina Bereta MarineTraffic, Athens, Greece

Elena Camossi NATO STO Centre for Maritime Research and Experimentation, La Spezia, Italy

Eva Chondrodima University of Piraeus, Piraeus, Greece

Gemma Galdon Clavell Eticas Research and Innovation, Barcelona, Spain

Jose Manuel Cordero CRIDA (Reference Center for Research, Development and Innovation in ATM), Madrid, Spain

Christos Doulkeridis University of Piraeus, Piraeus, Greece

Richard Dréo Arts et Metiers Institute of Technology, Ecole Navale, IRENav, Brest, France

Georg Fuchs Fraunhofer Institute IAIS, Sankt Augustin, Germany

Harris Georgiou University of Piraeus, Piraeus, Greece

Apostolis Glenis University of Piraeus, Piraeus, Greece

Ricardo Herranz Nommon Solutions and Technologies, Madrid, Spain

Clément Iphar NATO STO Centre for Maritime Research and Experimentation, La Spezia, Italy

Anne-Laure Jousset NATO STO Centre for Maritime Research and Experimentation, La Spezia, Italy

Nikolaos Koutroumanis University of Piraeus, Piraeus, Greece

Rodrigo Marcos Nommon Solutions and Technologies, Madrid, Spain

Panagiotis Nikitopoulos University of Piraeus, Piraeus, Greece

Kostas Patroumpas Athena Research Center, Marousi, Greece

Nikos Pelekis University of Piraeus, Piraeus, Greece

Petros Petrou University of Piraeus, Piraeus, Greece

Victoria Peuvrelle Eticas Research and Innovation, Barcelona, Spain

Manolis Pitsikalis Institute of Informatics & Telecommunications, NCSR Demokritos, Athens, Greece

Cyril Ray Arts et Metiers Institute of Technology, Ecole Navale, IRENav, Brest, France

Stefan Rüping Fraunhofer Institute IAIS, Sankt Augustin, Germany

Georgios M. Santipantakis University of Piraeus, Piraeus, Greece

David Scarlatti Boeing Research & Development Europe, Madrid, Spain

Stylianios Sideridis University of Piraeus, Piraeus, Greece

Panagiotis Tampakis University of Piraeus, Piraeus, Greece

Yannis Theodoridis University of Piraeus, Piraeus, Greece

Akrivi Vlachou University of Piraeus, Piraeus, Greece

Marios Vodas MarineTraffic, Athens, Greece

George A. Vouros University of Piraeus, Piraeus, Greece

Dimitris Zissis Department of Product & Systems Design Engineering, University of the Aegean, Lesbos, Greece
MarineTraffic, Athens, Greece

Maximilian Zocholl NATO STO Centre for Maritime Research and Experimentation, La Spezia, Italy

Acronyms

ADS-B	Automatic-dependent surveillance-broadcast
AI	Aircraft intent
AIS	Automatic identification system
ANFR	French Frequencies Agency
ANSP	Air Navigation Service Providers
API	Application programming interface
APM	Aircraft performance model
ATC	Air traffic control
ATCO	Air traffic controller
ATFM	Air traffic flow management
aTHS	approximate Trajectory Hot Spot
ATM	Air-traffic management
AtoN	Aids to navigation
BADA	Base of aircraft data
BGP	Basic graph pattern
CAS	Calibrated airspeed
CDO	Continuous descent operations
CEF	Complex event forecasting
CEP	Complex event processing
CER	Complex event recognition
CoG	Course over ground
COLREGs	Collision regulations
CP	Cutting point
CRS	Coordinate reference system
CSV	Comma separated values
CTOT	Calculated take off time
datAcron	Big data analytics for time-critical mobility forecasting
DBSCAN	Density-based spatial clustering of application with noise

DCB	Demand and capacity balancing
DG ENTR	European Commission's Directorate-General for Enterprise and Industry
DiStRDF	Distributed spatiotemporal RDF engine
DM	Data management/manager
DMP	Data management plan
DoF	Degrees of freedom
DSTs	Decision support tools
DTJ	Distributed subTrajectory Join
DTW	Dynamic time warping
EC	Entry count
EEA	European Environmental Agency
EEZ	Exclusive economic zone
EPIRB	Emergency Position Indicating Radio Beacon
ESRI	Environmental Systems Research Institute, Inc.
ETA	Estimated time of arrival
ETD	Estimated time of departure
ETOT	Estimated take off time
EUMSS	European Union Maritime Security Strategy
FAO	Food and Agriculture Organization
FDR	Flight recorded data
FLP	Future location predictor
FMS	Flight management system
FP	Flight plan
GFS	Global Forecast System
GLM	Generalized linear models
GPS	Global Positioning System
GSHHG	Global Self-consistent, Hierarchical, High-resolution Geography Database
HDFS	Hadoop distributed file system
HMM	Hidden Markov Model
ICAO	International Civil Aviation Organization
IFREMER	Institut français de recherche pour l'exploitation de la mer
IMO	International Maritime Organization
ISA	International Standard Atmosphere
ISR	Intelligence, surveillance, and reconnaissance
ITU	International Telecommunication Union
IUU	Illegal, unreported, and unregulated
IVA	Interactive visual analytics
JRC	European Commission Joint Research Centre
JSON	JavaScript Object Notation

KDE	Kernel density estimation
LCSS	Longest Common SubSequence
LD	Link discovery
LED	Low-level events detector
LNG	Liquid natural gas
LRIT	Long range tracking and identification
MBB	Minimum bounding box
MBR	Minimum bounding rectangle
METAR	Meteorological Terminal Aviation Routine Weather Report
METOC	Meteorological and oceanographic
MLlib	Machine learning library
MMSI	Maritime Mobile Service Identity
MOB	Man overboard
MSA	Maritime situation awareness
MSI	Maritime situational indicator
NGO	Non-governmental organization
NMEA	National Marine Electronics Association
NOAA	National Oceanic and Atmospheric Administration
ODC-By	Open Data Commons Attribution License
ODC-ODbL	Open Data Commons Open Database License
OGC	Open Geospatial Consortium
PbD	Privacy by design
PMM	Point-mass model
QAR	Quick access recorder
RBT	Reference business trajectory
RDF	Resource description framework
RFL	Requested flight level
RMSE	Root mean square error
ROC	Rate of climb
RoT	Rate of turn
RTEC	Event calculus for run-time reasoning
SAR	Search And Rescue
SBT	Shared business trajectory
SESAR	Single European Sky ATM Research
SI	Semantic integrator
SIGMET	Significant meteorological information
SoG	Speed over Ground
SSCR	Sum of similarity between cluster members and cluster representatives
STD	Semantic Trajectory Database
TAF	Terminal aerodrome forecast

TAS	True airspeed
TBO	Trajectory-based operations
TCL	Talis Community License
TDA	Trajectory data analytics
THS	Trajectory hot spot
TP	Trajectory predictors
TSA	Trajectory segmentation algorithm
TSS	Traffic separation scheme
TTL	Terse RDF Triple Language
UN	United Nations
VA	Visual analytics
VHF	Very high frequency
VMS	Vessel monitoring system
VTs	Vessel traffic system
WKT	Well-known text
WGS	World Geodetic System
WGS84	World Geodetic System 1984
WPI	World Port Index
YARN	Yet Another Resource Negotiator