Mobility Data

Mobility of people and goods is essential in the global economy. The ability to track the routes and patterns associated with this mobility offers unprecedented opportunities for developing new, smarter applications in different domains. Much of the current research is devoted to developing concepts, models, and tools to comprehend mobility data and make them manageable for these applications.

This book surveys the myriad facets of mobility data, from spatio-temporal data modeling, to data aggregation and warehousing, to data analysis, with a specific focus on monitoring people in motion (drivers, airplane passengers, crowds, and even animals in the wild). Written by a renowned group of worldwide experts, it presents a consistent framework that facilitates understanding of all these different facets, from basic definitions to state-of-the-art concepts and techniques, offering both researchers and professionals a thorough understanding of the applications and opportunities made possible by the development of mobility data.

Chiara Renso is a permanent researcher at the Institute of Information Science and Technologies at the Italian National Research Council, Italy. Her research interests are related to spatio-temporal data mining, reasoning, data mining query languages, semantic data mining, and trajectory data mining.

Stefano Spaccapietra is an honorary professor at the School of Computer and Communication Sciences, Swiss Federal Institute of Technology, in Lausanne, Switzerland, where he has been chairing the database laboratory for more than twenty years. Together with Christine Parent, he developed MADS, a conceptual spatio-temporal data model equipped with multi-representation support, which gained worldwide renown and has been used in several applications.

Esteban Zimányi is a professor and director of the Department of Computer & Decision Engineering of the Université Libre de Bruxelles. His current research interests include business intelligence, geographic information systems, spatio-temporal databases, data warehouses, and Semantic Web. He has co-authored two books and co-edited four books in the domains of spatio-temporal modeling, spatio-temporal data warehouses, and business intelligence.

Mobility Data

Modeling, Management, and Understanding

Edited by

CHIARA RENSO

ISTI Institute of National Research Council Italy

STEFANO SPACCAPIETRA

Ecole Polytechnique Federale de Lausanne

ESTEBAN ZIMÁNYI

Université Libre de Bruxelles



CAMBRIDGE UNIVERSITY PRESS

32 Avenue of the Americas, New York, NY 10013-2473, USA

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

> www.cambridge.org Information on this title: www.cambridge.org/9781107021716

> > C Cambridge University Press 2013

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2013

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication data

Mobility data : modeling, management, and understanding / [edited by] Chiara Renso, ISTI Institute of National Research Council (CNR), Pisa, Italy, Stefano Spaccapietra, Ecole Polytechnique Federale de Lausanne, Switzerland, Esteban Zimányi, Université Libre de Bruxelles, Belgium.

pages cm

Includes bibliographical references and index.

ISBN 978-1-107-02171-6 (hardback)

 Mobile computing. I. Renso, Chiara, 1968– editor of compilation. II. Spaccapietra, S., editor of compilation. III. Zimányi, Esteban, 1964– editor of compilation.

QA76.59.M725 2014

004.16-dc23 2013009544

ISBN 978-1-107-02171-6 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party Internet Web sites referred to in this publication and does not guarantee that any content on such Web sites is, or will remain, accurate or appropriate.

CONTENTS

List of Contributors	<i>page</i> xi
Preface	XV
Acknowledgments	xix

PART I MOBILITY DATA MODELING AND REPRESENTATION

1	Tra	jectories and Their Representations	3
	S. S _l	paccapietra, C. Parent, L. Spinsanti	
	1.1	Introduction	3
	1.2	Trajectory: Definition and Application Scenario	5
	1.3	From Raw Trajectories to Semantic Trajectories	10
	1.4	Trajectory Patterns and Behaviors	14
	1.5	Individual, Collective, and Sequence Trajectory Behavior	17
	1.6	Conclusions	20
	1.7	Bibliographic Notes	21
2	Tra	jectory Collection and Reconstruction	23
	G. Marketos, M.L. Damiani, N. Pelekis, Y. Theodoridis, Z. Yan		
	2.1	Introduction	23
	2.2	Tracking Trajectory Data	24
	2.3	Handling Trajectory Data	26
	2.4	Reconstructing Trajectories	30
	2.5	Protecting the Privacy of Individuals' Positions	34
	2.6	Conclusions	40
	2.7	Bibliographic Notes	40

v

vi	Contents	
3	Trajectory Databases	42
	R.H. Güting, T. Behr, C. Düntgen	
	3.1 Introduction	42
	3.2 Data Model and Query Language	45
	3.3 SECONDO	51
	3.4 Representations for Sets of Trajectories	56
	3.5 Indexing	58
	3.6 Hermes	59
	3.7 Conclusions	60
	3.8 Bibliographic Notes	60
4	Trajectory Data Warehouses	62
	A.A. Vaisman, E. Zimányi	
	4.1 Introduction	62
	4.2 Data Warehousing	63
	4.3 Running Example	65
	4.4 Querying Trajectory Data Warehouses	68
	4.5 Continuous Fields	73
	4.6 An Example Trajectory DW: GeoPKDD	76
	4.7 Conclusions	81
	4.8 Bibliographic Notes	81
5	Mobility and Uncertainty	83
	C. Silvestri, A.A. Vaisman	
	5.1 Introduction	83
	5.2 Causes of Uncertainty in Mobility Data	85
	5.3 Uncertainty Models for Spatio-Temporal Data	88
	5.4 Conclusions	100
	5.5 Bibliographic Notes	100
PA	RT II MOBILITY DATA UNDERSTANDING	
6	Mobility Data Mining	105
	M. Nanni	
	6.1 Introduction	105
	6.2 Local Trajectory Patterns/Behaviors	107
	6.3 Global Trajectory Models	113
	6.4 Conclusions	123
	6.5 Bibliographic Notes	125

	Contents	vii
7	Understanding Human Mobility Using Mobility Data	
	Mining	127
	C. Renso, R. Trasarti	
	7.1 The Mobility Knowledge Discovery Process	127
	7.2 The M-Atlas System	129
	7.3 Finding Behavior from Trajectory Data	140
	7.4 Conclusions	146
	7.5 Bibliographic Notes	147
8	Visual Analytics of Movement: A Rich Palette of	
	Techniques to Enable Understanding	149
	N. Andrienko, G. Andrienko	
	8.1 Introduction	149
	8.2 Looking at Trajectories	150
	8.3 Looking inside Trajectories: Attributes, Events, and Patterns	156
	8.4 Bird's Eye on Movement: Generalization and Aggregation	159
	8.5 Investigation of Movement in Context	167
	8.6 Conclusions	171
	8.7 Bibliographic Notes	171
9	Mobility Data and Privacy	174
	F. Giannotti, A. Monreale, D. Pedreschi	
	9.1 Introduction	174
	9.2 Basic Concepts for Data Privacy	175
	9.3 Privacy in Offline Mobility Data Analysis	178
	9.4 Privacy by Design in Data Mining	184
	9.5 Conclusions	192
	9.6 Bibliographic Notes	193
PA	RT III MOBILITY APPLICATIONS	
10	Car Traffic Monitoring	197
	D. Janssens, M. Nanni, S. Rinzivillo	
	10.1 Traffic Modeling and Transportation Science	197
	10.2 Data-Driven Traffic Models	199
	10.3 Data Understanding	201
	10.4 Analysis of Movement Behavior	206
	10.5 Conclusions	219
	10.6 Bibliographic Notes	220

viii	Contents	
11	Maritime Monitoring	221
	T. Devogele, L. Etienne, C. Ray	
	11.1 Maritime Context	221
	11.2 A Monitoring System Based on Data-Mining Processes	226
	11.3 Conclusions	238
	11.4 Bibliographic Notes	239
12	Air Traffic Analysis	240
	C. Hurter, G. Andrienko, N. Andrienko, R.H. Güting, M. Sakr	
	12.1 Introduction	240
	12.2 Motivation	241
	12.3 Data Set Description	243
	12.4 Direct Manipulation of Trajectories	244
	12.5 Event Extraction	249
	12.6 Complex Pattern Extraction Using a Moving Object Database	
	System	253
	12.7 Conclusions	257
	12.8 Bibliographic Notes	258
13	Animal Movement	259
	S. Focardi, F. Cagnacci	
	13.1 Introduction	259
	13.2 The Study of Animal Movement	264
	13.3 Conclusions	274
	13.4 Bibliographic Notes	275
14	Person Monitoring with Bluetooth Tracking	277
	M. Versichele, T. Neutens, N. Van de Weghe	
	14.1 The Difficult Nature of Measuring Human Mobility	277
	14.2 How Bluetooth Offers an Alternative Solution	279
	14.3 Case Studies	281
	14.4 Conclusions	292
	14.5 Bibliographic Notes	293
PA	RT IV FUTURE CHALLENGES AND CONCLUSIONS	
15	A Complexity Science Perspective on Human Mobility	297
	F. Giannotti, L. Pappalardo, D. Pedreschi, D. Wang	
	15.1 Models of Human Mobility	298
	15.2 Social Networks and Human Mobility	306
	15.3 Conclusions	312
	15.4 Bibliographic Notes	313

Contents	ix
16 Mobility and Geo-Social Networks	315
L. Spinsanti, M. Berlingerio, L. Pappalardo	
16.1 Introduction	315
16.2 Geo-Social Data and Mobility	317
16.3 Trajectory from Geo-Social Web	321
16.4 Geographic Information in Geo-Social Web	322
16.5 Open Issues	330
16.6 Conclusions	332
16.7 Bibliographic Notes	332
17 Conclusions	334
C. Renso, S. Spaccapietra, E. Zimányi	
Bibliography	341
Glossary	351
Author Index	357
Subject Index	358

Color plates section appears between pages 258 and 259

CONTRIBUTORS

Gennady Andrienko Fraunhofer Institute IAIS, Schloss Birlinghoven, Sankt-Augustin, Germany

Natalia Andrienko Fraunhofer Institute IAIS, Schloss Birlinghoven, Sankt-Augustin, Germany

Thomas Behr FernUniversität in Hagen, Hagen, Germany

Michele Berlingerio IBM Technology Campus (Building 3), Damastown Industrial Estate, Dublin, Ireland

Francesca Cagnacci Department of Biodiversity and Molecular Ecology, Research and Innovation Centre, Fondazione Edmund Mach, Trento, Italy

Maria Luisa Damiani Dipartimento di Informatica, Universita degli Studi di Milano, Milano, Italy

Thomas Devogele Laboratoire d'informatique, Université François Rabelais de Tours, Blois, France

Christian Düntgen FernUniversität in Hagen, Hagen, Germany

Laurent Etienne Department of Industrial Engineering, Dalhousie University, Halifax, NS, Canada

Stefano Focardi Istituto Dei Sistemi Complessi, ISC-CNR, Sesto Fiorentino, Italy

Fosca Giannotti KDD Lab, ISTI-CNR, Pisa, Italy

Ralf Hartmut Güting FernUniversität in Hagen, Hagen, Germany

Christophe Hurter Ecole Nationale de l'Aviation Civile (ENAC), Toulouse Cedex, France

xii

Contributors

Davy Janssens Transportation Research Institute (IMOB), Hasselt University, Diepenbeek, Belgium

Gerasimos Marketos Department of Informatics, University of Piraeus, Piraeus, Greece

Anna Monreale Computer Science Department, University of Pisa, Pisa, Italy

Mirco Nanni KDD Lab, ISTI-CNR, Pisa, Italy

Tijs Neutens Department of Geography, Ghent University, Gent, Belgium

Luca Pappalardo KDD Lab, ISTI-CNR, Pisa, Italy

Christine Parent ISI-HEC, Université de Lausanne, Lausanne, Switzerland

Dino Pedreschi KDD-Lab, Computer Science Department, University of Pisa, Pisa, Italy

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

Cyril Ray Naval Academy Research Institute, Brest Cedex 9, France

Chiara Renso KDD Lab, ISTI-CNR, Pisa, Italy

Salvatore Rinzivillo KDD Lab, ISTI-CNR, Pisa, Italy

Mahmoud Sakr FernUniversität in Hagen, Hagen, Germany

Claudio Silvestri Università Ca' Foscari di Venezia, Mestre (VE), Italy

Stefano Spaccapietra Faculté IC, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Laura Spinsanti JRC-TP262, Ispra (VA), Italy

Yannis Theodoridis Department of Informatics, University of Piraeus, Piraeus, Greece

Roberto Trasarti KDD Lab, ISTI-CNR, Pisa, Italy

Alejandro A. Vaisman Department of Computer and Decision Engineering (CoDE), Université Libre de Bruxelles, Bruxelles, Belgium

Nico Van de Weghe Department of Geography, Ghent University, Gent, Belgium

Mathias Versichele Department of Geography, Ghent University, Gent, Belgium

Contributors

xiii

Dashun Wang Center for Complex Network Research, Department of Physics, Northeastern University, Boston, MA, USA

Zhixian Yan Samsung R&D Center, San Jose, CA, USA

Esteban Zimányi Department of Computer and Decision Engineering (CoDE), Université Libre de Bruxelles, Bruxelles, Belgium

PREFACE

From the invention of the wheel to moon-landing rockets, technological progress over thousands of years has produced increasingly powerful and efficient transportation means, thus making moving easier and easier. Most recent progress in telecommunications has added new facets to mobility. We have now the ability to automatically keep track of our travel routes and even document them with information such as photos about the places we have been. This prompted the surge of small to huge databases holding mobility data, that is, the data about where and when we have been all over the world as well as during our daily trips to reach our workplace. Complementarily, more and more applications in a great variety of domains have been or are being developed to make intelligent use of mobility data.

While most of us are aware that our cellphones and cars equipped with a GPS facility do regularly generate signals conveying their geographical position (plus other data characterizing movement, e.g., acceleration and instant speed), not everybody is aware of what may happen later to this data, that is, how it can be used, by whom, and for what purpose. This book aims to introduce the potential answers to this question. The presentation of the material aims to make the book an easy read for all professionals (students included) in computer sciences and geoinformatics. Special attention has been given to show enough examples to optimize the understanding of the discussions. Moreover, application-oriented chapters have been included to illustrate a number of existing application domains that already benefit from using mobility data. All topics are covered to the level of detail that is compatible with a reasonable length of the book.

While the ultimate goal in mobility data processing is to solve high-level issues such as understanding how, when, where, and ultimately why objects (including persons and animals) move, elaborating the answer to these questions relies on a complex, multistep process, where the data sent by the data acquisition

xvi

Preface

device (e.g., a GPS/GSM device) are analyzed and transformed to be gradually turned into something readily meaningful for the targeted application. This process is sometimes referred to as the *Knowledge Discovery* (KD) process: from raw data to knowledge.

This book first offers an overview of the KD process as applied to mobility data. Each chapter from 1 to 8 discusses one of the issues involved.

Chapter 1 introduces the reader to the basic concepts and terms to deal with mobility data. Namely, the concept of trajectory is defined together with the various ways to approach this fundamental concept. Chapter 2 explains the most important techniques that can be used to collect the raw data from the acquisition devices and transform, homogenize, and prepare it for efficient use by applications, consistently with the application requirements. This includes potential modification of the raw data (e.g., anonymization and obfuscation) to meet privacy requirements. Chapter 3 focuses on how to store the mobility data in a database so that users can benefit from the existing know-how in database management. This is extremely important for this data to become operational with no delay. Chapter 4 similarly investigates the issues for storing mobility data in a data warehouse, opening to its use for decision-making applications interested in aggregated levels of knowledge rather than the detailed level of individual trajectories. Chapter 5 is specifically devoted to addressing the uncertainty issues that are inherent to mobility data, given that position measurements are affected by observational error and thus not necessarily as precise as applications would like them to be. The chapter closes the review of the basic data processing techniques needed for mobility data management.

With Chapter 6 the reader fully enters into the core of the knowledge management process (Part II of the book), that is, how to analyze the collected data to find its aggregated characteristics that can be of interest to the applications at hand. Movement patterns or trajectory behaviors are the core concern of the chapter. However, the lack of semantics of the extracted patterns makes the interpretation task far from obvious. To solve the mismatch following, Chapter 7 introduces the semantic dimension, thus closing the gap between the application quest for mobility information and the knowledge extracted from the data. The identification of semantic behaviors of the moving objects holds the final result of the KD process. Chapter 7 also presents a system, M-Atlas, which supports the whole mobility knowledge discovery process. Chapter 8 closes the knowledge extraction part of the book by showing through many illustrations how visualization of mobility data can be a very effective analysis tool to detect trends as well as singularities.

The last chapter in Part II of the book, Chapter 9, addresses the privacy issue, that is, how to ensure that mobility data do not violate the privacy regulations and constraints that aim at protecting individuals from the undue disclosure of personal data. This represents a very important concern as mobility data related

Preface

to moving persons can reveal details of the person's life that nobody wants to see exposed to public view. Moreover, as users of cellphones and computers we have very little control over what happens to the data that these electronic systems collect, most frequently without making us aware of the hidden data collection routines.

Part III of the book details a number of application examples that show the reader concrete uses of mobility data in a variety of domains. This part starts with the most frequently quoted application domain: car traffic. Obviously the popularity of this application domain is due to the relative ease of getting massive volumes of data from the GPS-equipped cars that have become available in recent years. Chapter 10 shows traffic application results from a variety of data repositories.

Chapter 11 is also devoted to traffic analyses, but its moving objects are boats and the moving space is the sea. This leads to a context that is quite different from cars moving in a city, as navigation rules and paths for boats are different from those of cars. The environmental data are quite different: cities show plenty of landmarks to which a human trajectory can be linked, and the same landmark (e.g., a commercial centre) can host a multiplicity of facilities that can be targeted by a moving person. Instead, the destination of a boat can usually be recognized without ambiguity, while its path is not arbitrary and has to avoid potentially hidden obstacles.

Chapter 12 closes the analysis of transportation means showing an air traffic control application, in which a variety of data sources, for example, meteorological data, have to be combined with trajectories of planes with very strong security constraints. The interesting feature of this application domain is its use of visualization tools that play an essential role in facilitating faster decision making.

Ecology is another very popular application domain that largely benefits from the availability of movement data. Chapter 13 discusses the evolution of scientific approaches to modeling animals' movement, from the formulation of the first hypotheses to modern mathematical models supporting statistical studies. It also discusses the devices that are used today for data acquisition of animals' movement.

The next application chapter, Chapter 14, covers aspects of human movement. Human movement has several unique features, such as unconstrained routes, unpredictability and sudden changes, variety of transportation means, and a richer variety of reasons for moving than animals have. In some contexts (e.g., large pedestrian crowds), traditional means of measuring mobility will not suffice for quantitative analyses. The chapter introduces the Bluetooth tracking methodology and some of its benefits in comparison with other methodologies. Despite the coarse nature of the data, exciting analyses such as crowd size estimations, flow analysis, pattern discovery, and profiling are possible.

xvii

xviii

Preface

Part IV concludes the book with three more chapters. The aim of this part is to introduce the newest developments that call for new forms and new uses of mobility data. Chapter 15 explores how the recent developments of network sciences can be applied to enriching mobility analysis approaches. This is a recent combination of scientific domains that together can significantly enhance our ability to understand movement. The second prospective chapter, Chapter 16, explores the peculiar forms of mobility data that can be gathered thanks to the popularity of social networks. Social network data is not necessarily in terms of trajectories, yet it implicitly conveys data about the movement of people. How to intelligently extract these data and analyze them is a new and exciting challenge. At last, the concluding Chapter 17 outlines some directions for future research in view of future applications. Obviously these are just a few examples; the real potential is huge.

ACKNOWLEDGMENTS

We would like to express our deepest gratitude to the many persons without whom this book would not exist.

We are obviously grateful to all the authors for their excellent contributions, as well as their availability for and commitment to writing and rewriting the several revisions of their chapters. We know how painful it may be to develop material that well harmonizes with the material provided by other authors, in particular authors from different disciplines and with different views. As editors we tried to help but also added our own constraints and guidelines, complementing those provided by the publisher. We do appreciate the cooperation efforts of the authors. We also are pleased to further thank those authors who accepted the extra task to review the first draft of a chapter written by other authors. They are listed below as internal reviewers.

We would like to stress the special role that Christine Parent generously accepted to play during all the steps of the process of developing the book. Thanks to her contributions, from many reviews to detailed suggestions for revising several chapters, as well as the glossary and the index, the final version of this book has reached the level of quality at which we were aiming.

We also owe special thanks to those colleagues who carefully reviewed the second version of the chapters, thus providing external insight and substantial comments to help the authors improve their chapters. Our gratitude is but a minimal compensation for their hard work. They are listed below as external reviewers.

We would also like to thank the European Project FP7-FET MODAP N. 245410 (http://www.modap.org/) and the COST Action MOVE N. IC0903 (http://www.move-cost.info/) for partially supporting the work of the authors and the editors of this book.

xх

Acknowledgments

Finally, we would like to warmly thank Lauren Cowles, from Cambridge University Press, for her continued support of this book. Since the day she came to us with the book proposal, her enthusiasm and encouragement throughout its writing helped significantly in giving us impetus to pursue our project to its end.

Internal Reviewers

- Gennady Andrienko
- Natalia Andrienko
- Christophe Claramunt
- Maria Luisa Damiani
- Thomas Devogele
- Fosca Giannotti
- Ralf Hartmut Güting
- Anna Monreale
- Mirco Nanni

- Christine Parent
- Nikos Pelekis
- Salvatore Rinzivillo
- Claudio Silvestri
- Laura Spinsanti
- Yannis Theodoridis
- Alejandro A. Vaisman
- Nico Van de Weghe
- **External Reviewers**
- Josep Domingo-Ferrer, Universitat Rovira i Virgili, Tarragona, Catalonia, Spain
- Eric Feron, School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, USA
- Georg Gartner, Institut für Geoinformation und Kartographie, Technischen Universität Wien, Austria
- Leticia Gómez, Instituto Tecnológico de Buenos Aires, Argentina
- Michael Goodchild, University of California, Santa Barbara, CA, USA
- Patrick Laube, Department of Geography, University of Zurich, Switzerland
- Michal May, Fraunhofer Institute for Intelligent Analysis and Informations Systems IAIS, Sankt Augustin, Germany
- Gavin McArdle, School of Computer Science and Informatics, University College Dublin, Ireland
- Rosa Meo, Dipartimento di Informatica, Università degli Studi di Torino, Italy
- Mohamed Mokbel, Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA
- Alejandro Pauly, Department of Computer & Information Science & Engineering, University of Florida, Gainesville, FL, USA
- Peter Smouse, Department of Ecology, Evolution & Natural Resources, Rutgers University, New Brunswick, NJ, USA
- Chaoming Song, Department of Physics, Northeastern University, Boston, MA, USA

Acknowledgments

xxi

- Kathleen Stewart, Department of Geography, The University of Iowa, Ames, IA, USA
- Jack van Wijk, Department of Mathematics and Computer Science, Eindhoven University of Technology
- Vassilios Verykios, Hellenic Open University, Greece