

Contents

Contributors	xv
About the Editors	xvii

1. Big Data and Transport Analytics: An Introduction

*Constantinos Antoniou, Loukas Dimitriou
and Francisco Câmara Pereira*

1 Introduction	1
2 Book Structure	3
Special Acknowledgments	4
References	5
Further Reading	5

Part I Methodological

2. Machine Learning Fundamentals

Francisco Câmara Pereira and Stanislav S. Borysov

1 Introduction	9
2 A Little Bit of History	11
3 Deep Neural Networks and Optimization	19
4 Bayesian Models	23
5 Basics of Machine Learning Experiments	24
6 Concluding Remarks	27
References	28
Further Reading	29

3. Using Semantic Signatures for Social Sensing in Urban Environments

*Krzysztof Janowicz, Grant McKenzie, Yingjie Hu, Rui Zhu
and Song Gao*

1 Introduction	31
2 Spatial Signatures	33
2.1 Spatial Point Pattern	34
2.2 Spatial Autocorrelations	34

2.3	Spatial Interactions With Other Geographic features	34
2.4	Place-Based Statistics	37
3	Temporal Signatures	37
4	Thematic Signatures	41
5	Examples	45
5.1	Comparing Place Types	45
5.2	Coreference Resolution Across Gazetteers	48
5.3	Geoprivacy	48
5.4	Temporally Enhanced Geolocation	49
5.5	Regional Variation	50
5.6	Extraction of Urban Functional Regions	50
6	Summary	52
	References	53
4.	Geographic Space as a Living Structure for Predicting Human Activities Using Big Data	
	<i>Bin Jiang and Zheng Ren</i>	
1	Introduction	55
2	Living Structure and the Topological Representation	57
3	Data and Data Processing	60
4	Prediction of Tweet Locations Through Living Structure	63
4.1	Correlations at the Scale of Thiessen Polygons	63
4.2	Correlations at the Scale of Natural Cities	64
4.3	Degrees of Wholeness or Life or Beauty	65
5	Implications on the Topological Representation and Living Structure	66
6	Conclusion	70
	Acknowledgments	70
	References	71
5.	Data Preparation	
	<i>Kristian Henrickson, Filipe Rodrigues and Francisco Camara Pereira</i>	
1	Introduction	73
2	Tools and Techniques	74
2.1	Scripting and Statistical Analysis Software	74
2.2	Database Management Software	76
2.3	Working With Web Data	79
3	Probe Vehicle Traffic Data	81
3.1	Formats and Protocols	81
3.2	Data Characteristics	83
3.3	Challenges	85
3.4	Data Preparation and Quality Control	88
4	Context Data	95
4.1	The Role of Context Data	95
4.2	Types of Context Data	96

4.3	Formats and Data Collection	99
4.4	Data Cleaning and Preparation	99
	References	102

6. Data Science and Data Visualization

Michalis Xyntarakis and Constantinos Antoniou

1	Introduction	107
2	Structured Visualization	115
3	Multidimensional Data Visualization Techniques	120
3.1	Parallel Coordinates	121
3.2	Multidimensional Scaling (MDS)	123
3.3	t-Distributed Stochastic Neighbor Embedding for High-Dimensional Data Sets (t-SNE)	123
4	Case Studies	124
4.1	Experimental Setup	124
4.2	Car Characteristics Data Set	125
4.3	Congestion on I95	128
4.4	Dimensionality Reduction on NYC Taxi Flows	132
4.5	Dimensionality Reduction on the NYC Turnstile Data Set	140
5	Conclusions	142
	References	143
	Further Reading	144

7. Model-Based Machine Learning for Transportation

Inon Peled, Filipe Rodrigues and Francisco Câmara Pereira

1	Introduction	145
1.1	Background Concepts	146
1.2	Notation	147
2	Case Study 1: Taxi Demand in New York City	147
2.1	Initial Probabilistic Model: Linear Regression	147
2.2	Key Components of MBML	148
2.3	Inference	150
2.4	Model Improvements	152
3	Case Study 2: Travel Mode Choices	155
3.1	Improvement: Hierarchical Modeling	157
4	Case Study 3: Freeway Occupancy in San Francisco	160
4.1	Autoregressive Model	160
4.2	State-Space Model	160
4.3	Linear Dynamical Systems	161
4.4	Common Enhancements to LDS	162
4.5	NonLinear Variations on LDS	164
5	Case Study 4: Incident Duration Prediction	165
5.1	Preprocessing	166
5.2	Bag-of-Words Encoding	166
5.3	Latent Dirichlet Allocation	166

6	Summary	169
6.1	Further Reading	169
	References	170
8.	Textual Data in Transportation Research: Techniques and Opportunities	
	<i>Aseem Kinra, Samaneh Beheshti Kashi, Francisco Câmara Pereira, François Combes and Werner Rothengatter</i>	
1	Introduction	173
2	Big Textual Data, Text Sources, and Text Mining	174
2.1	Meaning of Text in the Context of Computational Linguistics	174
2.2	Text Mining	176
2.3	Text Mining Process Model	177
2.4	Textual Data Sources in Transportation	179
3	Fundamental Concepts and Techniques in Literature	181
3.1	Topic Modeling	183
3.2	Word2Vec—Text Embeddings With Deep Learning	185
4	Application Examples of Big Textual Data in Transportation	187
4.1	Developing Transportation and Logistics Performance Classifiers Using NLTK and Naïve Bayes	187
4.2	Understanding the Public Opinion Toward Driverless Cars With Topic Modeling	190
4.3	Predicting Taxi Demand in Special Events With Text Embeddings and Deep Learning	192
5	Conclusions	194
	References	195
	Further Reading	197

Part II

Applications

9.	Statewide Comparison of Origin-Destination Matrices Between California Travel Model and Twitter	
	<i>Jae Hyun Lee, Adam Davis, Elizabeth McBride and Konstantinos G. Goulias</i>	
1	Introduction	201
2	California Statewide Travel Demand Model	203
3	Twitter Data	204
4	Trip Extraction Methods	206
5	Models for Matrix Conversion	208
5.1	Tobit Regression Model	211
5.2	Latent Class Regression Model	212
6	Summary and Conclusion	225
	References	226

10. Transit Data Analytics for Planning, Monitoring, Control, and Information

Haris N. Koutsopoulos, Zhenliang Ma, Peyman Noursalehi and Yiwen Zhu

1	Introduction	229
2	Measuring System Performance From the Passenger's Point of View	232
2.1	The Individual Reliability Buffer Time (IRBT)	233
2.2	Denied Boarding	238
3	Decision Support With Predictive Analytics	243
3.1	Framework	245
3.2	Application: Provision of Crowding Predictive Information	250
4	Optimal Design of Transit Demand Management Strategies	252
4.1	Framework and Problem Formulation	254
4.2	Application: Prepeak Discount Design	256
5	Conclusion	258
	Acknowledgments	259
	References	259
	Further Reading	261

11. Data-Driven Traffic Simulation Models: Mobility Patterns Using Machine Learning Techniques

Vasileia Papathanasopoulou, Constantinos Antoniou and Haris N. Koutsopoulos

1	New Modeling Challenges and Data Opportunities	263
1.1	New Modeling Requirements	264
1.2	New Data Sources	264
1.3	Future Challenges	265
2	Background	265
3	Data-Driven Traffic Performance Modeling: Overall Framework	267
3.1	Modeling Approach	267
3.2	Model Components	268
4	Application to Mesoscopic Modeling	276
4.1	Data and Experimental Design	276
4.2	Case Study Setup	276
4.3	Application and Results	277
5	Application to Microscopic Traffic Modeling	277
5.1	Data and Experimental Design	278
5.2	Case Study Setup	279
5.3	Application and Results	279
6	Application to Weak Lane Discipline Modeling	280
6.1	Data and Experimental Design	281
6.2	Case Study Setup	282
6.3	Application and Results	284
7	Network-Wide Application	287
7.1	Implementation Aspects	287

7.2 Case Study Setup	289	3.3 Patch Detection and Measurement Using Support Vector Machines (SVM)	374
7.3 Results	289	4 Conclusions	377
8 Conclusions	290	References	379
Acknowledgments	291		
References	291		
12. Big Data and Road Safety: A Comprehensive Review		15. Collaborative Positioning for Urban Intelligent Transportation Systems (ITS) and Personal Mobility (PM): Challenges and Perspectives	
<i>Katerina Stylianou, Loukas Dimitriou and Mohamed Abdel-Aty</i>		<i>Vassilis Gikas, Guenther Retscher and Allison Kealy</i>	
1 Introduction	297	1 Introduction	382
2 The Role of Big Data in Traffic Safety Analysis	298	2 C-ITS in Support of the Smart Cities Concept	383
2.1 Real-Time Crash Prediction	299	2.1 Scientific and Policy Perspectives of Urban C-ITS	383
2.2 Driving Behavior	327	2.2 Taxonomy of Urban C-ITS Applications	386
3 ADAS and Autonomous Vehicles (AVs)	332	3 User Requirements for Urban C-ITS	387
4 Conclusions	336	3.1 Requirements Overview	387
References	337	3.2 Positioning Requirements and Parameters Definition	387
		4 Positioning Technologies for Urban ITS	389
13. A Back-Engineering Approach to Explore Human Mobility Patterns Across Megacities Using Online Traffic Maps		4.1 Radio Frequency-Based (RF) Technologies	393
<i>Vana Gkania and Loukas Dimitriou</i>		4.2 MEMS-Based Inertial Navigation	397
1 Introduction	345	4.3 Optical Technologies	398
2 Data and Traffic Information Extraction Methods	347	5 Measuring Types and Positioning Techniques	399
2.1 Cities Characteristics	347	5.1 Absolute Positioning Techniques	399
2.2 Data Gathering and Preprocessing	349	5.2 Relative and Hybrid Positioning Techniques	401
2.3 Extracting Traffic Information by Image Processing	349	6 CP for C-ITS	402
3 Temporal and Spatiotemporal Mobility Patterns	351	6.1 From Single Sensor Positioning to CP	402
3.1 Temporal Patterns	352	6.2 Fusion Algorithms and Techniques for CP	403
3.2 Spatiotemporal Patterns	356	7 Application Cases of Integrated Urban C-ITS	404
4 Dynamic Clustering and Propagation of Congestion	358	7.1 Case 1: Smart-Bike Systems as a Component of Urban C-ITS	404
5 Conclusions	362	7.2 Case 2: Smart Intersection for Traffic Control and Safety	406
References	363	8 Discussion, Perspectives, and Conclusions	407
		References	409
14. Pavement Patch Defects Detection and Classification Using Smartphones, Vibration Signals and Video Images		Further Reading	413
<i>Symeon E. Christodoulou, Charalambos Kyriakou and George Hadjidemetriou</i>			
1 Introduction	365	Conclusions	415
2 Brief Literature Review	367	Index	419
2.1 Vibration-Based Methods	367		
2.2 Vision-Based Methods	368		
3 Methodology	369		
3.1 Anomaly Detection Using ANNs and Timeseries Analysis of Vibration Signals	369		
3.2 Anomaly Detection Using Entropic-Filter Image Segmentation	371		