

# Contents

Preface xiii

Acknowledgments xxi

## I Concepts and Framework 1

### 1 Overview 3

1.1 Introduction 3

1.2 Definition of Urban Computing 4

1.3 General Framework 5

1.3.1 *Brief and Example* 5

1.3.2 *Functions of Each Layer* 5

1.4 Key Urban-Computing Challenges 10

1.4.1 *Urban-Sensing Challenges* 10

1.4.2 *Urban Data Management Challenges* 15

1.4.3 *Urban Data Analytics Challenges* 17

1.4.4 *Urban Service Challenges* 22

1.5 Urban Data 28

1.5.1 *Taxonomy of Urban Data* 28

1.5.2 *Geographical Data* 30

1.5.3 *Traffic Data on Road Networks* 31

1.5.4 *Mobile Phone Data* 33

1.5.5 *Commuting Data* 34

1.5.6 *Environmental-Monitoring Data* 35

1.5.7 *Social Network Data* 37

1.5.8 *Energy* 38

1.5.9 *Economy* 39

1.5.10 *Health Care* 39

1.6 Public Datasets 39

References 41

<b>2</b>	<b>Urban-Computing Applications</b>	<b>49</b>
2.1	Introduction	49
2.2	Urban Computing for Urban Planning	49
2.2.1	<i>Gleaning Underlying Problems in Transportation Networks</i>	49
2.2.2	<i>Discovering Functional Regions</i>	52
2.2.3	<i>Detecting a City's Boundaries</i>	54
2.2.4	<i>Facility and Resource Deployment</i>	55
2.3	Urban Computing for Transportation Systems	58
2.3.1	<i>Improving Driving Experiences</i>	58
2.3.2	<i>Improving Taxi Services</i>	61
2.3.3	<i>Improving Bus Services</i>	65
2.3.4	<i>Subway Services</i>	66
2.3.5	<i>Bike-Sharing Systems</i>	67
2.4	Urban Computing for the Environment	70
2.4.1	<i>Air Quality</i>	71
2.4.2	<i>Noise Pollution</i>	75
2.4.3	<i>Urban Water</i>	78
2.5	Urban Computing for Urban Energy Consumption	81
2.5.1	<i>Gas Consumption</i>	81
2.5.2	<i>Electricity Consumption</i>	83
2.6	Urban Computing for Social Applications	84
2.6.1	<i>Concepts of Location-Based Social Networks</i>	84
2.6.2	<i>Understanding Users in Location-Based Social Networks</i>	85
2.6.3	<i>Location Recommendations</i>	85
2.7	Urban Computing for the Economy	88
2.7.1	<i>Location Selection for Businesses</i>	88
2.7.2	<i>Optimizing Urban Logistics</i>	89
2.8	Urban Computing for Public Safety and Security	91
2.8.1	<i>Detecting Urban Anomalies</i>	91
2.8.2	<i>Predicting the Flow of Crowds</i>	95
2.9	Summary	96
	References	97
<b>II</b>	<b>Urban Sensing and Data Acquisition</b>	<b>113</b>
<b>3</b>	<b>Urban Sensing</b>	<b>115</b>
3.1	Introduction	115
3.1.1	<i>Four Paradigms of Urban Sensing</i>	115
3.1.2	<i>General Framework of Urban Sensing</i>	119
3.2	Sensor and Facility Deployment	121
3.2.1	<i>Finding Optimal Meeting Points</i>	121
3.2.2	<i>Maximizing Coverage</i>	125

3.2.3	<i>Learning-to-Rank Candidates</i>	129
3.2.4	<i>Minimizing Uncertainty</i>	132
3.3	Human-Centric Urban Sensing	133
3.3.1	<i>Data Evaluation</i>	134
3.3.2	<i>Participant Recruitment and Task Design</i>	136
3.4	Filling Missing Values	139
3.4.1	<i>Problem and Challenges</i>	139
3.4.2	<i>Spatial Models</i>	141
3.4.3	<i>Temporal Models</i>	145
3.4.4	<i>Spatiotemporal Models</i>	148
3.5	Summary	150
	References	150
<b>III</b>	<b>Urban Data Management</b>	<b>155</b>
<b>4</b>	<b>Spatiotemporal Data Management</b>	<b>157</b>
4.1	Introduction	157
4.1.1	<i>Data Structures</i>	157
4.1.2	<i>Queries</i>	158
4.1.3	<i>Indexes</i>	158
4.1.4	<i>Retrieval Algorithms</i>	161
4.2	Data Structures	161
4.2.1	<i>Point-Based Spatial Static Data</i>	161
4.2.2	<i>Point-Based Spatial Time Series Data</i>	163
4.2.3	<i>Point-Based Spatiotemporal Data</i>	163
4.2.4	<i>Network-Based Spatial Static Data</i>	164
4.2.5	<i>Network-Based Spatial Time Series Data</i>	165
4.2.6	<i>Network-Based Spatiotemporal Data</i>	165
4.3	Spatial Data Management	167
4.3.1	<i>Grid-Based Spatial Index</i>	168
4.3.2	<i>Quadtree-Based Spatial Index</i>	169
4.3.3	<i>K-D Tree-Based Spatial Index</i>	172
4.3.4	<i>R-Tree-Based Spatial Index</i>	175
4.4	Spatiotemporal Data Management	180
4.4.1	<i>Managing Spatial Static and Temporal Dynamic Data</i>	180
4.4.2	<i>Moving-Object Databases</i>	183
4.4.3	<i>Trajectory Data Management</i>	189
4.5	Hybrid Indexes for Managing Multiple Datasets	198
4.5.1	<i>Queries and Motivations</i>	198
4.5.2	<i>Spatial Key Words</i>	200
4.5.3	<i>Indexes for Managing Multiple Datasets</i>	210
4.6	Summary	213
	References	214

<b>5</b>	<b>Introduction to Cloud Computing</b>	<b>219</b>
5.1	Introduction	219
5.2	Storage	221
5.2.1	<i>SQL Databases</i>	221
5.2.2	<i>Azure Storage</i>	223
5.2.3	<i>Redis Cache</i>	233
5.3	Computing	235
5.3.1	<i>Virtual Machine</i>	236
5.3.2	<i>Cloud Services</i>	237
5.3.3	<i>HDInsight</i>	240
5.4	Applications	255
5.4.1	<i>Web Apps</i>	256
5.4.2	<i>Mobile Apps</i>	257
5.4.3	<i>API Apps</i>	258
5.5	Summary	259
	References	259
<b>6</b>	<b>Managing Spatiotemporal Data in the Cloud</b>	<b>263</b>
6.1	Introduction	263
6.1.1	<i>Challenges</i>	263
6.1.2	<i>General Data Management Schemes on the Cloud</i>	265
6.2	Managing Point-Based Data	267
6.2.1	<i>Managing Point-Based Spatiotemporal Static Data</i>	267
6.2.2	<i>Managing Point-Based Spatial Static and Temporal Dynamic Data</i>	272
6.2.3	<i>Managing Point-Based Spatiotemporal Dynamic Data</i>	278
6.3	Managing Network-Based Data	284
6.3.1	<i>Managing Spatiotemporal Static Networks</i>	284
6.3.2	<i>Managing Network-Based Spatial Static and Temporally Dynamic Data</i>	290
6.3.3	<i>Managing Network-Based Spatiotemporal Dynamic Data</i>	294
6.4	Urban Big-Data Platform	301
6.5	Summary	303
<b>IV</b>	<b>Urban Data Analytics</b>	<b>305</b>
<b>7</b>	<b>Fundamental Data-Mining Techniques for Urban Data</b>	<b>307</b>
7.1	Introduction	307
7.1.1	<i>General Framework of Data Mining</i>	307
7.1.2	<i>Relationship between Data Mining and Related Technologies</i>	309
7.2	Data Preprocessing	314
7.2.1	<i>Data Cleaning</i>	314
7.2.2	<i>Data Transformation</i>	317
7.2.3	<i>Data Integration</i>	319

7.3	Frequent Pattern Mining and Association Rules	326
7.3.1	<i>Basic Concepts</i>	327
7.3.2	<i>Frequent Itemset Mining Methods</i>	329
7.3.3	<i>Sequential Pattern Mining</i>	334
7.3.4	<i>Frequent Subgraph Pattern Mining</i>	341
7.4	Clustering	345
7.4.1	<i>Concepts</i>	345
7.4.2	<i>Partitioning Clustering Methods</i>	347
7.4.3	<i>Density-Based Clustering</i>	349
7.4.4	<i>Hierarchical Clustering Methods</i>	354
7.5	Classification	356
7.5.1	<i>Concepts</i>	356
7.5.2	<i>Naïve Bayesian Classification</i>	359
7.5.3	<i>Decision Trees</i>	360
7.5.4	<i>Support Vector Machines</i>	364
7.5.5	<i>Classification with Imbalanced Data</i>	366
7.6	Regression	370
7.6.1	<i>Linear Regression</i>	370
7.6.2	<i>Autoregression</i>	373
7.6.3	<i>Regression Tree</i>	374
7.7	Outlier and Anomaly Detection	376
7.7.1	<i>Proximity-Based Outlier Detection</i>	377
7.7.2	<i>Statistic-Based Outlier Detection</i>	380
7.8	Summary	384
	References	385
<b>8</b>	<b>Advanced Machine-Learning Techniques for Spatiotemporal Data</b>	<b>391</b>
8.1	Introduction	391
8.2	Unique Properties of Spatiotemporal Data	392
8.2.1	<i>Spatial Properties of Spatiotemporal Data</i>	392
8.2.2	<i>Temporal Properties</i>	394
8.3	Collaborative Filtering	396
8.3.1	<i>Basic Models: User Based and Item Based</i>	396
8.3.2	<i>Collaborative Filtering for Spatiotemporal Data</i>	399
8.4	Matrix Factorization	404
8.4.1	<i>Basic Matrix Factorization Methods</i>	404
8.4.2	<i>Matrix Factorization for Spatiotemporal Data</i>	406
8.5	Tensor Decomposition	411
8.5.1	<i>Basic Concepts of Tensors</i>	411
8.5.2	<i>Methods of Tensor Decomposition</i>	413
8.5.3	<i>Tensor Decomposition for Spatiotemporal Data</i>	416
8.6	Probabilistic Graphical Models	421
8.6.1	<i>General Concepts</i>	421
8.6.2	<i>Bayesian Networks</i>	423

8.6.3	<i>Markov Random Field</i>	433
8.6.4	<i>Bayesian Networks for Spatiotemporal Data</i>	433
8.6.5	<i>Markov Networks for Spatiotemporal Data</i>	448
8.7	Deep Learning	454
8.7.1	<i>Artificial Neural Networks</i>	455
8.7.2	<i>Convolutional Neural Networks</i>	459
8.7.3	<i>Recurrent Neural Networks</i>	464
8.7.4	<i>Deep Learning for Spatiotemporal Data</i>	467
8.8	Reinforcement Learning	471
8.8.1	<i>Concepts of Reinforcement Learning</i>	471
8.8.2	<i>Tabular Action-Value Methods</i>	474
8.8.3	<i>Approximate Methods</i>	481
8.9	Summary	484
	References	485

## **9 Cross-Domain Knowledge Fusion 491**

9.1	Introduction	491
9.1.1	<i>Relationship to Traditional Data Integration</i>	493
9.1.2	<i>Relationship to Heterogeneous Information Networks</i>	494
9.2	Stage-Based Knowledge Fusion	495
9.3	Feature-Based Knowledge Fusion	498
9.3.1	<i>Feature Concatenation with Regularization</i>	498
9.3.2	<i>Deep Learning-Based Knowledge Fusion</i>	502
9.4	Semantic Meaning-Based Knowledge Fusion	504
9.4.1	<i>Multi-View-Based Knowledge Fusion</i>	505
9.4.2	<i>Similarity-Based Knowledge Fusion</i>	511
9.4.3	<i>Probabilistic Dependency-Based Knowledge Fusion</i>	517
9.4.4	<i>Transfer Learning-Based Knowledge Fusion</i>	518
9.5	Comparison between Different Fusion Methods	526
9.5.1	<i>Volume, Properties, and Insight of Datasets</i>	527
9.5.2	<i>The Goal of a Machine-Learning Task</i>	528
9.5.3	<i>Learning Approach of Machine-Learning Algorithms</i>	528
9.5.4	<i>Efficiency and Scalability</i>	528
9.6	Summary	529
	References	530

## **10 Advanced Topics in Urban Data Analytics 535**

10.1	How to Select Useful Datasets	535
10.1.1	<i>Understanding Target Problems</i>	536
10.1.2	<i>Insights behind Data</i>	537
10.1.3	<i>Validating Assumptions</i>	538
10.2	Trajectory Data Mining	542
10.2.1	<i>Trajectory Data</i>	545
10.2.2	<i>Trajectory Preprocessing</i>	546

10.2.3	<i>Trajectory Data Management</i>	556
10.2.4	<i>Uncertainty in a Trajectory</i>	556
10.2.5	<i>Trajectory Pattern Mining</i>	560
10.2.6	<i>Trajectory Classification</i>	566
10.2.7	<i>Anomalies Detection from Trajectories</i>	568
10.2.8	<i>Transferring Trajectories to Other Representations</i>	570
10.3	<i>Combining Machine Learning with Data Management</i>	576
10.3.1	<i>Motivation</i>	576
10.3.2	<i>Boosting Machine Learning with Indexing Structures</i>	580
10.3.3	<i>Scale Down Candidates for Machine Learning</i>	584
10.3.4	<i>Derive Bounds to Prune Computing Spaces for Machine Learning</i>	585
10.4	<i>Interactive Visual Data Analytics</i>	587
10.4.1	<i>Incorporating Multiple Complex Factors</i>	587
10.4.2	<i>Adjusting Parameters without Prior Knowledge</i>	587
10.4.3	<i>Drilling Down into Results</i>	588
10.5	<i>Summary</i>	588
	<i>References</i>	589
	<i>About the Author</i>	599
	<i>Index</i>	601