

Contents

Part I Basic Data Science

1	Introduction: Data Science and BigData Computing	3
	Li M. Chen	
1.1	Data Mining and Cloud Computing: The Prelude of BigData and Data Science	3
1.2	BigData Era	4
1.3	The Meaning of Data Sciences	6
1.4	Problems Related to Data Science	7
1.5	Mathematical Problems in Data Science	9
1.6	Mathematics, Data Science, and Data Scientists in Industry	11
1.7	Remark: Discussion on the Future Problems in Data Science	13
	References	14
2	Overview of Basic Methods for Data Science	17
	Li M. Chen	
2.1	“Hardware” and “Software” of Data Science	17
2.1.1	Searching and Optimization	18
2.1.2	Decision Making	18
2.1.3	Classification	18
2.1.4	Learning	19
2.2	Graph-Theoretic Methods	20
2.2.1	Review of Graphs	20
2.2.2	Breadth First Search and Depth First Search	21
2.2.3	Dijkstra’s Algorithm for the Shortest Path	22
2.2.4	Minimum Spanning Tree	22
2.3	Statistical Methods	23
2.4	Classification, Clustering, and Pattern Recognition	25
2.4.1	k -Nearest Neighbor Method	25
2.4.2	k -Means Method	25
2.5	Numerical Methods and Data Reconstruction in Science and Engineering	26

2.6	Algorithm Design and Computational Complexity	28
2.6.1	Concepts of Data Structures and Databases	28
2.6.2	Queues, Stacks, and Linked Lists	29
2.6.3	Quadtrees, Octrees, and <i>R</i> -trees	29
2.6.4	NP-Hard Problems and Approximation of Solutions	31
2.7	Online Searching and Matching	32
2.7.1	Google Search	32
2.7.2	Matching	34
2.7.3	High Dimensional Search	34
2.8	Remarks: Relationship Among Data Science, Database, Networking, and Artificial Intelligence	35
	References	35
3	Relationship and Connectivity of Incomplete Data Collection	39
	Li M. Chen	
3.1	Current Challenges of Problems in Data Science	39
3.1.1	Relations and Connectedness	41
3.1.2	Multiple Scaling and Multiple Level Sampling	41
3.1.3	Relationship and Connectivity Among Data	42
3.2	Generalized λ -Connectedness	42
3.2.1	λ -Connectedness on Undirected Graphs	43
3.2.2	λ -Connectedness on Directed Graphs	44
3.2.3	Potential Function ρ and the Measure μ	45
3.3	λ -Connected Decomposition and Image Segmentation	46
3.3.1	λ -Connected Region Growing Segmentation	46
3.3.2	λ -Connected Split-and-Merge Segmentation	47
3.4	λ -Connectedness for Data Reconstruction	49
3.4.1	λ -Connected Fitting	50
3.4.2	Intelligent Data Fitting	50
3.4.3	λ -Connected Fitting with Smoothness	51
3.5	Maximum Connectivity Spanning Tree and λ -Value	53
3.6	λ -Connectedness and Topological Data Analysis	55
3.7	Remark: Future Concerns	56
	References	57

Part II Data Science Problems and Machine Learning

4	Machine Learning for Data Science: Mathematical or Computational	63
	Li M. Chen	
4.1	Decision Trees and Boosting Process	63
4.2	Neural Networks	64
4.3	Genetic Algorithms	66
4.4	Functional and Variational Learning	67
4.5	Support Vector Machine Algorithms	68
4.6	Computational Learning Theory	70

- 4.7 Remarks: Statistical Learning Algorithms and BigData..... 71
- References..... 73
- 5 Images, Videos, and BigData 75**
- Li M. Chen
- 5.1 Images and Videos in BigData Times 75
- 5.2 Concepts of Image Processing: Filtering,
Segmentation, and Recognition..... 77
 - 5.2.1 Image Filtering..... 77
 - 5.2.2 Segmentation 79
 - 5.2.3 Image Recognition..... 79
- 5.3 The Five Philosophies of Image Segmentation 80
- 5.4 Image Segmentation Methods 82
 - 5.4.1 The Maximum Entropy Method and the
Minimum Variance Method 82
 - 5.4.2 Learning in λ -Connectedness with Maximum
Entropy 83
 - 5.4.3 λ -Connected Segmentation and the Minimum
Variance Method..... 85
 - 5.4.4 λ -Connectedness and the Mumford-Shah
Method 87
- 5.5 Graph-Cut Based Image Segmentation and Laplacian Matrices ... 87
- 5.6 Segmentation, BigData, and Subspace Clustering 89
 - 5.6.1 Quadtree and Octree Based Image Segmentation 89
 - 5.6.2 General Subspace Clustering..... 92
- 5.7 Object Tracking in Video Images..... 93
 - 5.7.1 Videos, Compression, and Video Storing Formats 93
 - 5.7.2 Object Tracking..... 94
 - 5.7.3 Tracking Algorithms 95
- 5.8 Future Concerns: BigData Related Image Segmentation..... 98
- References..... 99
- 6 Topological Data Analysis 101**
- Li M. Chen
- 6.1 Why Topology for Data Sets? 101
- 6.2 Concepts: Cloud Data, Decomposition, Simplex,
Complex, and Topology 102
- 6.3 Algorithmic Geometry and Topology 103
 - 6.3.1 Algorithms for Delaunay Triangulations
and Voronoi Diagrams..... 104
 - 6.3.2 Manifold Learning on Cloud Data..... 105
- 6.4 Persistent Homology and Data Analysis 109
 - 6.4.1 Euler Characteristics and Homology Groups 109
 - 6.4.2 Data Analysis Using Topology 112

6.5	Digital Topology Methods and Fast Implementation	113
6.5.1	2D Digital Holes and Betti Numbers	113
6.5.2	3D Genus Computation	115
6.6	New Developments in Persistent Homology and Analysis	117
6.6.1	Coverage of Sensor Networks Using Persistent Homology	117
6.6.2	Statistics and Machine Learning Combined with Topological Analysis	118
6.7	Remarks: Topological Computing and Applications in the Future	118
6.7.1	λ -Connectedness and Topological Data Analysis	119
6.7.2	Hierarchy of λ -Connectedness and Topological Data Analysis	120
6.7.3	Topological Computing in Cloud Computers	121
	References	123
7	Monte Carlo Methods and Their Applications in Big Data Analysis .	125
	Hao Ji and Yaohang Li	
7.1	Introduction	125
7.2	The Basic of Monte Carlo	126
7.3	Variance Reduction	128
7.3.1	Stratified Sampling	128
7.3.2	Control Variates	129
7.3.3	Antithetic Variates	129
7.3.4	Importance Sampling	130
7.4	Examples of Monte Carlo Methods in Data Science	131
7.4.1	Case Study 1: Estimation of Sum	131
7.4.2	Case Study 2: Monte Carlo Linear Solver	132
7.4.3	Case Study 3: Image Recovery	133
7.4.4	Case Study 4: Matrix Multiplication	134
7.4.5	Case Study 5: Low-Rank Approximation	136
7.5	Summary	138
	References	139
 Part III Selected Topics in Data Science		
8	Feature Extraction via Vector Bundle Learning	143
	Risheng Liu and Zhixun Su	
8.1	Introduction	143
8.2	Vector Bundle Learning	144
8.2.1	Mathematical Fundamentals	144
8.2.2	Vector Bundle Learning Framework	145
8.2.3	General Framework for Feature Extraction	147
8.3	Vector Bundle Discriminant Analysis	148

8.4	Experimental Results.....	150
8.4.1	Face Recognition	150
8.4.2	Handwriting Digits Classification	151
8.4.3	Discussions	152
8.5	Unifying Previous Algorithms in VBL Framework	153
8.5.1	Local Algorithms	153
8.5.2	Global Algorithms	154
8.6	Conclusions.....	156
	References.....	156
9	Curve Interpolation and Financial Curve Construction	159
	Pengfei Huang, Haiyan Wang, Ping Wu, and Yifei Li	
9.1	Introduction.....	159
9.2	Basic Interpolation Methods for Interest Rate Curves.....	160
9.2.1	Linear Interpolation	160
9.2.2	Cubic Spline Interpolation	161
9.2.3	Monotone and Convex Splines.....	162
9.3	Rational Cubic Interpolation Method for Positivity Preservation ..	163
9.4	Reconstruction of the Market Interest Rate Data	166
9.5	Analysis of the Interest Curve Constructed by Rational Cubic Interpolation.....	167
9.6	Conclusion.....	169
	References.....	169
10	Advanced Methods in Variational Learning: Segmentation with Intensity Inhomogeneity	171
	Jack Spencer and Ke Chen	
10.1	Introduction.....	172
10.2	Image Segmentation.....	172
10.3	Variational Methods	174
10.4	Variant Mumford–Shah Model [8]	176
10.4.1	Convergence Behaviour of VMS	177
10.5	Modified Models	178
10.5.1	Method 1	179
10.5.2	Method 2	179
10.6	Results	181
10.7	Conclusions.....	186
	References.....	186
11	An On-Line Strategy of Groups Evacuation from a Convex Region in the Plane	189
	Bo Jiang, Yuan Liu, Hao Zhang, and Xuehou Tan	
11.1	Introduction.....	189
11.2	Preliminaries.....	190
11.3	The Multi-Source Points Evacuation Problem.....	191

- 11.4 Concluding Remarks 197
- References 198
- 12 A New Computational Model of Bigdata 201**
 - Binhai Zhu
 - 12.1 Introduction 201
 - 12.2 MSM: The Master/Slave Multiprocessor Model 204
 - 12.3 Processing Large Data sets with an MSM 205
 - 12.3.1 Selection 206
 - 12.3.2 Diameter in any Fixed Dimension 207
 - 12.3.3 Sorting and Element Uniqueness 207
 - 12.3.4 Relaxing the Master Memory Constraint in an MSM 208
 - 12.4 Concluding Remarks 209
 - References 209
- Index 211**



<http://www.springer.com/978-3-319-25125-7>

Mathematical Problems in Data Science

Theoretical and Practical Methods

Chen, L.M.; Su, Z.; Jiang, B.

2015, XV, 213 p. 64 illus., 42 illus. in color., Hardcover

ISBN: 978-3-319-25125-7