

# Contents

<b>1</b>	<b>Introduction</b>	1
1.1	On Spatial Data Mining and Knowledge Discovery	1
1.2	What Makes Spatial Data Mining Different	2
1.3	On Spatial Knowledge	3
1.4	On Spatial Data	4
1.5	Basic Tasks of Knowledge Discovery in Spatial Data	5
1.6	Issues of Knowledge Discovery in Spatial Data	10
1.7	Methodological Background for Knowledge Discovery in Spatial Data	11
1.8	Organization of the Book	12
<b>2</b>	<b>Discovery of Intrinsic Clustering in Spatial Data</b>	13
2.1	A Brief Background About Clustering	13
2.2	Discovery of Clustering in Space by Scale Space Filtering	17
2.2.1	On Scale Space Theory for Hierarchical Clustering	18
2.2.2	Hierarchical Clustering in Scale Space	20
2.2.3	Cluster Validity Check	25
2.2.4	Clustering Selection Rules	29
2.2.5	Some Numerical Examples	31
2.2.6	Discovering Land Covers in Remotely Sensed Images	32
2.2.7	Mining of Seismic Belts in Vector-Based Databases	36
2.2.8	Visualization of Temporal Seismic Activities via Scale Space Filtering	42
2.2.9	Summarizing Remarks on Clustering by Scale Space Filtering	46
2.3	Partitioning of Spatial Data by a Robust Fuzzy Relational Data Clustering Method	49
2.3.1	On Noise and Scale in Spatial Partitioning	50
2.3.2	Clustering Algorithm with Multiple Scale Parameters for Noisy Data	51
2.3.3	Robust Fuzzy Relational Data Clustering Algorithm	54

- 2.3.4 Numerical Experiments ..... 57
- 2.4 Partitioning of Spatial Object Data by Unidimensional Scaling ..... 61
  - 2.4.1 A Note on the Use of Unidimensional Scaling ..... 61
  - 2.4.2 Basic Principle of Unidimensional Scaling in Data Clustering ..... 62
  - 2.4.3 Analysis of Simulated Data ..... 64
  - 2.4.4 UDS Clustering of Remotely Sensed Data ..... 66
- 2.5 Unraveling Spatial Objects with Arbitrary Shapes Through Mixture Decomposition Clustering ..... 70
  - 2.5.1 On Noise and Mixture Distributions in Spatial Data ..... 70
  - 2.5.2 A Remark on the Mining of Spatial Features with Arbitrary Shapes ..... 74
  - 2.5.3 A Spatial-Feature Mining Model (RFMM) Based on Regression-Class Mixture Decomposition (RCMD) ..... 75
  - 2.5.4 The RFMM with Genetic Algorithm (RFMM-GA) ..... 78
  - 2.5.5 Applications of RFMM-GA in the Mining of Features in Remotely Sensed Images ..... 80
- 2.6 Cluster Characterization by the Concept of Convex Hull ..... 84
  - 2.6.1 A Note on Convex Hull and its Computation ..... 84
  - 2.6.2 Basics of the Convex Hull Computing Neural Network (CHCNN) Model ..... 86
  - 2.6.3 The CHCNN Architecture ..... 89
  - 2.6.4 Applications in Cluster Characterization ..... 94
- 3 Statistical Approach to the Identification of Separation Surface for Spatial Data ..... 97**
  - 3.1 A Brief Background About Statistical Classification ..... 97
  - 3.2 The Bayesian Approach to Data Classification ..... 100
    - 3.2.1 A Brief Description of Bayesian Classification Theory ..... 100
    - 3.2.2 Naive Bayes Method and Feature Selection in Data Classification ..... 101
    - 3.2.3 The Application of Naïve Bayes Discriminant Analysis in Client Segmentation for Product Marketing ..... 102
    - 3.2.4 Robust Bayesian Classification Model ..... 112
  - 3.3 Mixture Discriminant Analysis ..... 113
    - 3.3.1 A Brief Statement About Mixture Discriminant Analysis ..... 113
    - 3.3.2 Mixture Discriminant Analysis by Optimal Scoring ..... 114
    - 3.3.3 Analysis Results and Interpretations ..... 115
  - 3.4 The Logistic Model for Data Classification ..... 117
    - 3.4.1 A Brief Note About Using Logistic Regression as a Classifier ..... 117
    - 3.4.2 Data Manipulation for Client Segmentation ..... 118
    - 3.4.3 Logistic Regression Models and Strategies for Credit Card Promotion ..... 119
    - 3.4.4 Model Comparisons and Validations ..... 125

- 3.5 Support Vector Machine for Spatial Classification ..... 130
  - 3.5.1 Support Vector Machine as a Classifier ..... 130
  - 3.5.2 Basics of Support Vector Machine ..... 131
  - 3.5.3 Experiments on Feature Extraction and Classification  
by SVM ..... 136
- 4 Algorithmic Approach to the Identification of Classification**
  - Rules or Separation Surface for Spatial Data ..... 143**
  - 4.1 A Brief Background About Algorithmic Classification ..... 143
  - 4.2 The Classification Tree Approach to the Discovery of Classification  
Rules in Data ..... 145
    - 4.2.1 A Brief Description of Classification and Regression tree  
(CART) ..... 145
    - 4.2.2 Client Segmentation by CART ..... 148
  - 4.3 The Neural Network Approach to the Classification  
of Spatial Data ..... 156
    - 4.3.1 On the Use of Neural Networks in Spatial Classification ..... 156
    - 4.3.2 The Knowledge-Integrated Radial Basis Function (RBF)  
Model for Spatial Classification ..... 159
    - 4.3.3 An Elliptical Basis Function Network for Spatial  
Classification ..... 172
  - 4.4 Genetic Algorithms for Fuzzy Spatial Classification Systems ..... 183
    - 4.4.1 A Brief Note on Using GA to Discover Fuzzy  
Classification Rules ..... 183
    - 4.4.2 A General Framework of the Fuzzy Classification System ..... 184
    - 4.4.3 Fuzzy Rule Acquisition by GANGO ..... 186
    - 4.4.4 An Application in the Classification of Remote  
Sensing Data ..... 194
  - 4.5 The Rough Set Approach to the Discovery of Classification  
Rules in Spatial Data ..... 196
    - 4.5.1 Basic Ideas of the Rough Set Methodology for Knowledge  
Discovery ..... 196
    - 4.5.2 Basic Notions Related to Spatial Information Systems  
and Rough Sets ..... 198
    - 4.5.3 Interval-Valued Information Systems and Data  
Transformation ..... 200
    - 4.5.4 Knowledge Discovery in Interval-Valued Information  
Systems ..... 202
    - 4.5.5 Discovery of Classification Rules for Remotely  
Sensed Data ..... 205
    - 4.5.6 Classification of Tree Species with Hyperspectral Data ..... 214
  - 4.6 A Vision-Based Approach to Spatial Classification ..... 216
    - 4.6.1 On Scale and Noise in Spatial Data Classification ..... 216
    - 4.6.2 The Vision-Based Classification Method ..... 218
    - 4.6.3 Experimental Results ..... 219

- 4.7 A Remark on the Choice of Classifiers ..... 221
- 5 Discovery of Spatial Relationships in Spatial Data ..... 223**
  - 5.1 On Mining Spatial Relationships in Spatial Data ..... 223
  - 5.2 Discovery of Local Patterns of Spatial Association ..... 225
    - 5.2.1 On the Measure of Local Variations of Spatial Associations ... 225
    - 5.2.2 Local Statistics and their Expressions as a Ratio  
of Quadratic Forms ..... 227
  - 5.3 Discovery of Spatial Non-Stationarity Based on  
the Geographically Weighted Regression Model ..... 236
    - 5.3.1 On Modeling Spatial Non-Stationarity within the  
Parameter-Varying Regression Framework ..... 236
    - 5.3.2 Geographically Weighted Regression and the  
Local–Global Issue About Spatial Non-Stationarity ..... 238
    - 5.3.3 Local Variations of Regional Industrialization in  
Jiangsu Province, P.R. China ..... 244
    - 5.3.4 Discovering Spatial Pattern of Influence of Extreme  
Temperatures on Mean Temperatures in China ..... 250
  - 5.4 Testing for Spatial Autocorrelation in Geographically  
Weighted Regression ..... 254
  - 5.5 A Note on the Extensions of the GWR Model ..... 258
  - 5.6 Discovery of Spatial Non-Stationarity Based on  
the Regression-Class Mixture Decomposition Method ..... 260
    - 5.6.1 On Mixture Modeling of Spatial Non-Stationarity  
in a Noisy Environment ..... 260
    - 5.6.2 The Notion of a Regression Class ..... 262
    - 5.6.3 The Discovery of Regression Classes under Noise  
Contamination ..... 263
    - 5.6.4 The Regression-Class Mixture Decomposition (RCMD)  
Method for knowledge Discovery in Mixed Distribution ..... 267
    - 5.6.5 Numerical Results and Observations ..... 271
    - 5.6.6 Comments About the RCMD Method ..... 272
    - 5.6.7 A Remote Sensing Application ..... 275
    - 5.6.8 An Overall View about the RCMD Method ..... 276
- 6 Discovery of Structures and Processes in Temporal Data ..... 277**
  - 6.1 A Note on the Discovery of Generating Structures or  
Processes of Time Series Data ..... 277
  - 6.2 The Wavelet Approach to the Mining of Scaling  
Phenomena in Time Series Data ..... 279
    - 6.2.1 A Brief Note on Wavelet Transform ..... 279
    - 6.2.2 Basic Notions of Wavelet Analysis ..... 280
    - 6.2.3 Wavelet Transforms in High Dimensions ..... 285
    - 6.2.4 Other Data Mining Tasks by Wavelet Transforms ..... 286
    - 6.2.5 Wavelet Analysis of Runoff Changes in the Middle  
and Upper Reaches of the Yellow River in China ..... 286

- 6.2.6 Wavelet Analysis of Runoff Changes of the Yangtze River Basin ..... 289
- 6.3 Discovery of Generating Structures of Temporal Data with Long-Range Dependence ..... 292
  - 6.3.1 A Brief Note on Multiple Scaling and Intermittency of Temporal Data ..... 292
  - 6.3.2 Multifractal Approach to the Identification of Intermittency in Time Series Data ..... 293
  - 6.3.3 Experimental Study on Intermittency of Air Quality Data Series ..... 297
- 6.4 Finding the Measure Representation of Time Series with Intermittency ..... 301
  - 6.4.1 Multiplicative Cascade as a Characterization of the Time Series Data ..... 301
  - 6.4.2 Experimental Results ..... 302
- 6.5 Discovery of Spatial Variability in Time Series Data ..... 307
  - 6.5.1 Multifractal Analysis of Spatial Variability Over Time ..... 307
  - 6.5.2 Detection of Spatial Variability of Rainfall Intensity ..... 309
- 6.6 Identification of Multifractality and Spatio-Temporal Long Range Dependence in Multiscaling Remote Sensing ..... 312
  - 6.6.1 A Note on Multifractality and Long-Range Dependence in Remote Sensing Data ..... 312
  - 6.6.2 A Proposed Methodology for the Analysis of Multifractality and Long-Range Dependence in Remote Sensing Data ..... 314
- 6.7 A Note on the Effect of Trends on the Scaling Behavior of Time Series with Long-Range Dependence ..... 317
- 7 Summary and Outlooks ..... 321**
  - 7.1 Summary ..... 321
  - 7.2 Directions for Further Research ..... 322
    - 7.2.1 Discovery of Hierarchical Knowledge Structure from Relational Spatial Data ..... 322
    - 7.2.2 Errors in Spatial Knowledge Discovery ..... 324
    - 7.2.3 Other Challenges ..... 326
  - 7.3 Concluding Remark ..... 327
- Bibliography ..... 329**
- Author Index ..... 351**
- Subject Index ..... 357**



<http://www.springer.com/978-3-642-02663-8>

Knowledge Discovery in Spatial Data

Leung, Y.

2009, XXIX, 360 p. 113 illus., Hardcover

ISBN: 978-3-642-02663-8