
Contents

1 Introduction to Knowledge Discovery and Data Mining	
<i>Oded Maimon, Lior Rokach</i>	1

Part I Preprocessing Methods

2 Data Cleansing: A Prelude to Knowledge Discovery	
<i>Jonathan I. Maletic, Andrian Marcus</i>	19

3 Handling Missing Attribute Values	
<i>Jerzy W. Grzymala-Busse, Witold J. Grzymala-Busse</i>	33

4 Geometric Methods for Feature Extraction and Dimensional Reduction - A Guided Tour	
<i>Christopher J.C. Burges</i>	53

5 Dimension Reduction and Feature Selection	
<i>Barak Chizi, Oded Maimon</i>	83

6 Discretization Methods	
<i>Ying Yang, Geoffrey I. Webb, Xindong Wu</i>	101

7 Outlier Detection	
<i>Irad Ben-Gal</i>	117

Part II Supervised Methods

8 Supervised Learning	
<i>Lior Rokach, Oded Maimon</i>	133

9 Classification Trees	
<i>Lior Rokach, Oded Maimon</i>	149

10 Bayesian Networks

Paola Sebastiani, Maria M. Abad, Marco F. Ramoni 175

11 Data Mining within a Regression Framework

Richard A. Berk 209

12 Support Vector Machines

Armin Shmilovici 231

13 Rule Induction

Jerzy W. Grzymala-Busse 249

Part III Unsupervised Methods

14 A survey of Clustering Algorithms

Lior Rokach 269

15 Association Rules

Frank Höppner 299

16 Frequent Set Mining

Bart Goethals 321

17 Constraint-based Data Mining

Jean-Francois Boulicaut, Baptiste Jeudy 339

18 Link Analysis

Steve Donoho 355

Part IV Soft Computing Methods

19 A Review of Evolutionary Algorithms for Data Mining

Alex A. Freitas 371

20 A Review of Reinforcement Learning Methods

Oded Maimon, Shahar Cohen 401

21 Neural Networks For Data Mining

G. Peter Zhang 419

22 Granular Computing and Rough Sets - An Incremental Development

Tsau Young ('T. Y.') Lin, Churn-Jung Liau 445

23 Pattern Clustering Using a Swarm Intelligence Approach

Swagatam Das, Ajith Abraham 469

24 Using Fuzzy Logic in Data Mining	
<i>Lior Rokach</i>	505

Part V Supporting Methods

25 Statistical Methods for Data Mining	
<i>Yoav Benjamini, Moshe Leshno</i>	523

26 Logics for Data Mining	
<i>Petr Hájek</i>	541

27 Wavelet Methods in Data Mining	
<i>Tao Li, Sheng Ma, Mitsunori Ogihara</i>	553

28 Fractal Mining - Self Similarity-based Clustering and its Applications	
<i>Daniel Barbara, Ping Chen</i>	573

29 Visual Analysis of Sequences Using Fractal Geometry	
<i>Noa Ruschin Rimini, Oded Maimon</i>	591

30 Interestingness Measures - On Determining What Is Interesting	
<i>Sigal Sahar</i>	603

31 Quality Assessment Approaches in Data Mining	
<i>Maria Halkidi, Michalis Vazirgiannis</i>	613

32 Data Mining Model Comparison	
<i>Paolo Giudici</i>	641

33 Data Mining Query Languages	
<i>Jean-Francois Boulicaut, Cyrille Masson</i>	655

Part VI Advanced Methods

34 Mining Multi-label Data	
<i>Grigorios Tsoumakas, Ioannis Katakis, Ioannis Vlahavas</i>	667

35 Privacy in Data Mining	
<i>Vicenç Torra</i>	687

36 Meta-Learning - Concepts and Techniques	
<i>Ricardo Vilalta, Christophe Giraud-Carrier, Pavel Brazdil</i>	717

37 Bias vs Variance Decomposition For Regression and Classification	
<i>Pierre Geurts</i>	733

38 Mining with Rare Cases	
<i>Gary M. Weiss</i>	747
39 Data Stream Mining	
<i>Mohamed Medhat Gaber, Arkady Zaslavsky, Shonali Krishnaswamy</i>	759
40 Mining Concept-Drifting Data Streams	
<i>Haixun Wang, Philip S. Yu, Jiawei Han</i>	789
41 Mining High-Dimensional Data	
<i>Wei Wang, Jiong Yang</i>	803
42 Text Mining and Information Extraction	
<i>Moty Ben-Dov, Ronen Feldman</i>	809
43 Spatial Data Mining	
<i>Shashi Shekhar, Pusheng Zhang, Yan Huang</i>	837
44 Spatio-temporal clustering	
<i>Slava Kisilevich, Florian Mansmann, Mirco Nanni, Salvatore Rinzivillo</i>	855
45 Data Mining for Imbalanced Datasets: An Overview	
<i>Nitesh V. Chawla</i>	875
46 Relational Data Mining	
<i>Sašo Džeroski</i>	887
47 Web Mining	
<i>Johannes Fürnkranz</i>	913
48 A Review of Web Document Clustering Approaches	
<i>Nora Oikonomakou, Michalis Vazirgiannis</i>	931
49 Causal Discovery	
<i>Hong Yao, Cory J. Butz, Howard J. Hamilton</i>	949
50 Ensemble Methods in Supervised Learning	
<i>Lior Rokach</i>	959
51 Data Mining using Decomposition Methods	
<i>Lior Rokach, Oded Maimon</i>	981
52 Information Fusion - Methods and Aggregation Operators	
<i>Vicenç Torra</i>	999
53 Parallel And Grid-Based Data Mining – Algorithms, Models and Systems for High-Performance KDD	
<i>Antonio Congiusta, Domenico Talia, Paolo Trunfio</i>	1009

54 Collaborative Data Mining	
<i>Steve Moyle</i>	1029
55 Organizational Data Mining	
<i>Hamid R. Nemati, Christopher D. Barko</i>	1041
56 Mining Time Series Data	
<i>Chotirat Ann Ratanamahatana, Jessica Lin, Dimitrios Gunopulos, Eamonn Keogh, Michail Vlachos, Gautam Das</i>	1049
<hr/>	
Part VII Applications	
<hr/>	
57 Multimedia Data Mining	
<i>Zhongfei (Mark) Zhang, Ruofei Zhang</i>	1081
58 Data Mining in Medicine	
<i>Nada Lavrač, Blaž Zupan</i>	1111
59 Learning Information Patterns in Biological Databases - Stochastic Data Mining	
<i>Gautam B. Singh</i>	1137
60 Data Mining for Financial Applications	
<i>Boris Kovalerchuk, Evgenii Vityaev</i>	1153
61 Data Mining for Intrusion Detection	
<i>Anoop Singhal, Sushil Jajodia</i>	1171
62 Data Mining for CRM	
<i>Kurt Thearling</i>	1181
63 Data Mining for Target Marketing	
<i>Nissan Levin, Jacob Zahavi</i>	1189
64 NHECD - Nano Health and Environmental Commented Database	
<i>Oded Maimon, Abel Browarnik</i>	1221
<hr/>	
Part VIII Software	
<hr/>	
65 Commercial Data Mining Software	
<i>Qingyu Zhang, Richard S. Segall</i>	1245
66 Weka-A Machine Learning Workbench for Data Mining	
<i>Eibe Frank, Mark Hall, Geoffrey Holmes, Richard Kirkby, Bernhard Pfahringer, Ian H. Witten, Len Trigg</i>	1269
Index	1279