
Contents

List of Tables	xi
List of Figures	xiii
Preface	xv
1 Introduction	1
1.1 Defining the Area	1
1.2 The Content and the Organization of This Book	4
1.3 The Audience of This Book	6
1.4 Further Readings	6
I Models	9
2 Co-Clustering	11
2.1 Introduction	11
2.2 Related Work	12
2.3 Model Formulation and Analysis	13
2.3.1 Block Value Decomposition	13
2.3.2 NBVD Method	17
3 Heterogeneous Relational Data Clustering	21
3.1 Introduction	21
3.2 Related Work	23
3.3 Relation Summary Network Model	24
4 Homogeneous Relational Data Clustering	29
4.1 Introduction	29
4.2 Related Work	32
4.3 Community Learning by Graph Approximation	33
5 General Relational Data Clustering	39
5.1 Introduction	39
5.2 Related Work	40
5.3 Mixed Membership Relational Clustering	42
5.4 Spectral Relational Clustering	45

6	Multiple-View Relational Data Clustering	47
6.1	Introduction	47
6.2	Related Work	49
6.3	Background and Model Formulation	50
6.3.1	A General Model for Multiple-View Unsupervised Learning	51
6.3.2	Two Specific Models: Multiple-View Clustering and Multiple-View Spectral Embedding	53
7	Evolutionary Data Clustering	57
7.1	Introduction	57
7.2	Related Work	59
7.3	Dirichlet Process Mixture Chain (DPChain)	60
7.3.1	DPChain Representation	61
7.4	HDP Evolutionary Clustering Model (HDP-EVO)	63
7.4.1	HDP-EVO Representation	63
7.4.2	Two-Level CRP for HDP-EVO	65
7.5	Infinite Hierarchical Hidden Markov State Model	66
7.5.1	iH ² MS Representation	67
7.5.2	Extention of iH ² MS	68
7.5.3	Maximum Likelihood Estimation of HTM	69
7.6	HDP Incorporated with HTM (HDP-HTM)	70
7.6.1	Model Representation	70
II	Algorithms	73
8	Co-Clustering	75
8.1	Nonnegative Block Value Decomposition Algorithm	75
8.2	Proof of the Correctness of the NBVD Algorithm	78
9	Heterogeneous Relational Data Clustering	83
9.1	Relation Summary Network Algorithm	83
9.2	A Unified View to Clustering Approaches	90
9.2.1	Bipartite Spectral Graph Partitioning	90
9.2.2	Binary Data Clustering with Feature Reduction	90
9.2.3	Information-Theoretic Co-Clustering	91
9.2.4	K-Means Clustering	92
10	Homogeneous Relational Data Clustering	95
10.1	Hard CLGA Algorithm	95
10.2	Soft CLGA Algorithm	97
10.3	Balanced CLGA Algorithm	101

11 General Relational Data Clustering	105
11.1 Mixed Membership Relational Clustering Algorithm	105
11.1.1 MMRC with Exponential Families	105
11.1.2 Monte Carlo E-Step	108
11.1.3 M-Step	109
11.1.4 Hard MMRC Algorithm	112
11.2 Spectral Relational Clustering Algorithm	114
11.3 A Unified View to Clustering	118
11.3.1 Semi-Supervised Clustering	118
11.3.2 Co-Clustering	119
11.3.3 Graph Clustering	120
12 Multiple-View Relational Data Clustering	123
12.1 Algorithm Derivation	123
12.1.1 Multiple-View Clustering Algorithm	124
12.1.2 Multiple-View Spectral Embedding Algorithm	127
12.2 Extensions and Discussions	129
12.2.1 Evolutionary Clustering	129
12.2.2 Unsupervised Learning with Side Information	130
13 Evolutionary Data Clustering	133
13.1 DPChain Inference	133
13.2 HDP-EVO Inference	134
13.3 HDP-HTM Inference	136
III Applications	139
14 Co-Clustering	141
14.1 Data Sets and Implementation Details	141
14.2 Evaluation Metrics	142
14.3 Results and Discussion	143
15 Heterogeneous Relational Data Clustering	147
15.1 Data Sets and Parameter Setting	147
15.2 Results and Discussion	150
16 Homogeneous Relational Data Clustering	153
16.1 Data Sets and Parameter Setting	153
16.2 Results and Discussion	155
17 General Relational Data Clustering	159
17.1 Graph Clustering	159
17.2 Bi-clustering and Tri-Clustering	161
17.3 A Case Study on Actor-Movie Data	163
17.4 Spectral Relational Clustering Applications	164
17.4.1 Clustering on Bi-Type Relational Data	164

17.4.2 Clustering on Tri-Type Relational Data	166
18 Multiple-View and Evolutionary Data Clustering	169
18.1 Multiple-View Clustering	169
18.1.1 Synthetic Data	169
18.1.2 Real Data	172
18.2 Multiple-View Spectral Embedding	173
18.3 Semi-Supervised Clustering	174
18.4 Evolutionary Clustering	175
IV Summary	179
References	185
Index	195