
Contents

Part I Background Introduction

1	Broad Learning Introduction	3
1.1	What Is Broad Learning	3
1.2	Problems and Challenges of Broad Learning	4
1.2.1	Cross-Source Information Fusion	4
1.2.2	Cross-Source Knowledge Discovery	6
1.2.3	Challenges of Broad Learning	6
1.3	Comparison with Other Learning Tasks	7
1.3.1	Broad Learning vs. Deep Learning	7
1.3.2	Broad Learning vs. Ensemble Learning	8
1.3.3	Broad Learning vs. Transfer Learning vs. Multi-Task Learning	8
1.3.4	Broad Learning vs. Multi-View, Multi-Source, Multi-Modal, Multi-Domain Learning	9
1.4	Book Organization	10
1.4.1	Part I	10
1.4.2	Part II	11
1.4.3	Part III	11
1.4.4	Part IV	12
1.5	Who Should Read This Book	12
1.6	How to Read This Book	13
1.6.1	To Readers	13
1.6.2	To Instructors	13
1.6.3	Supporting Materials	14
1.7	Summary	14
1.8	Bibliography Notes	14
1.9	Exercises	15
	References	15
2	Machine Learning Overview	19
2.1	Overview	19
2.2	Data Overview	20
2.2.1	Data Types	20
2.2.2	Data Characteristics	26
2.2.3	Data Pre-processing and Transformation	26

2.3	Supervised Learning: Classification	32
2.3.1	Classification Learning Task and Settings	33
2.3.2	Decision Tree	34
2.3.3	Support Vector Machine	40
2.4	Supervised Learning: Regression	46
2.4.1	Regression Learning Task	46
2.4.2	Linear Regression	46
2.4.3	Lasso	47
2.4.4	Ridge	48
2.5	Unsupervised Learning: Clustering	48
2.5.1	Clustering Task	49
2.5.2	K-Means	50
2.5.3	DBSCAN	52
2.5.4	Mixture-of-Gaussian Soft Clustering	54
2.6	Artificial Neural Network and Deep Learning	56
2.6.1	Artificial Neural Network Overview	57
2.6.2	Deep Learning	62
2.7	Evaluation Metrics	66
2.7.1	Classification Evaluation Metrics	66
2.7.2	Regression Evaluation Metrics	68
2.7.3	Clustering Evaluation Metrics	69
2.8	Summary	71
2.9	Bibliography Notes	72
2.10	Exercises	72
	References	73
3	Social Network Overview	77
3.1	Overview	77
3.2	Graph Essentials	78
3.2.1	Graph Representations	78
3.2.2	Connectivity in Graphs	80
3.3	Network Measures	82
3.3.1	Degree	82
3.3.2	Centrality	85
3.3.3	Closeness	91
3.3.4	Transitivity and Social Balance	96
3.4	Network Categories	98
3.4.1	Homogeneous Network	99
3.4.2	Heterogeneous Network	102
3.4.3	Aligned Heterogeneous Networks	106
3.5	Meta Path	111
3.5.1	Network Schema	111
3.5.2	Meta Path in Heterogeneous Social Networks	111
3.5.3	Meta Path Across Aligned Heterogeneous Social Networks	114
3.5.4	Meta Path-Based Network Measures	116
3.6	Network Models	117
3.6.1	Random Graph Model	118
3.6.2	Preferential Attachment Model	120

3.7	Summary	121
3.8	Bibliography Notes	122
3.9	Exercises	123
	References	124

Part II Information Fusion: Social Network Alignment

4	Supervised Network Alignment	129
4.1	Overview	129
4.2	Supervised Network Alignment Problem Definition	130
4.3	Supervised Full Network Alignment	131
4.3.1	Feature Extraction for Anchor Links	132
4.3.2	Supervised Anchor Link Prediction Model	138
4.3.3	Stable Matching	139
4.4	Supervised Partial Network Alignment	142
4.4.1	Partial Network Alignment Description	142
4.4.2	Inter-Network Meta Path Based Feature Extraction	143
4.4.3	Class-Imbalance Classification Model	146
4.4.4	Generic Stable Matching	149
4.5	Anchor Link Inference with Cardinality Constraint	151
4.5.1	Loss Function for Anchor Link Prediction	151
4.5.2	Cardinality Constraint Description	153
4.5.3	Joint Optimization Function	154
4.5.4	Problem and Algorithm Analysis	156
4.5.5	Distributed Algorithm	158
4.6	Summary	160
4.7	Bibliography Notes	160
4.8	Exercises	161
	References	163
5	Unsupervised Network Alignment	165
5.1	Overview	165
5.2	Heuristics Based Unsupervised Network Alignment	166
5.2.1	User Names Based Network Alignment Heuristics	166
5.2.2	Profile Based Network Alignment Heuristics	169
5.3	Pairwise Homogeneous Network Alignment	170
5.3.1	Heuristics Based Network Alignment Model	171
5.3.2	IsoRank	173
5.3.3	IsoRankN	175
5.3.4	Matrix Inference Based Network Alignment	176
5.4	Multiple Homogeneous Network Alignment with Transitivity Penalty	177
5.4.1	Multiple Network Alignment Problem Description	178
5.4.2	Unsupervised Multiple Network Alignment	179
5.4.3	Transitive Network Matching	184
5.5	Heterogeneous Network Co-alignment	186
5.5.1	Network Co-alignment Problem Description	187
5.5.2	Anchor Link Co-inference	188
5.5.3	Network Co-matching	195

5.6	Summary	198
5.7	Bibliography Notes	198
5.8	Exercises	199
	References	200
6	Semi-supervised Network Alignment	203
6.1	Overview	203
6.2	Semi-supervised Learning: Overview	204
6.2.1	Semi-supervised Learning Problem Setting	204
6.2.2	Semi-supervised Learning Models	205
6.2.3	Active Learning	209
6.2.4	Positive and Unlabeled (PU) Learning	211
6.3	Semi-supervised Network Alignment	212
6.3.1	Loss Function for Labeled and Unlabeled Instances	212
6.3.2	Cardinality Constraint on Anchor Links	214
6.3.3	Joint Objective Function for Semi-supervised Network Alignment	215
6.4	Active Network Alignment	215
6.4.1	Anchor Link Label Query Strategy	216
6.4.2	Active Network Alignment Objective Function	219
6.5	Positive and Unlabeled (PU) Network Alignment	221
6.5.1	PU Network Alignment Problem Formulation and Preliminary	221
6.5.2	PU Network Alignment Model	222
6.6	Summary	224
6.7	Bibliography Notes	224
6.8	Exercises	225
	References	225
Part III Broad Learning: Knowledge Discovery Across Aligned Networks		
7	Link Prediction	229
7.1	Overview	229
7.2	Traditional Single Homogeneous Network Link Prediction	230
7.2.1	Unsupervised Link Prediction	230
7.2.2	Supervised Link Prediction	232
7.2.3	Matrix Factorization Based Link Prediction	233
7.3	Heterogeneous Network Collective Link Prediction	235
7.3.1	Introduction to LBSNs	235
7.3.2	Collective Link Prediction	236
7.3.3	Information Accumulation and Feature Extraction	236
7.3.4	Collective Link Prediction Model	239
7.4	Cold Start Link Prediction for New Users	242
7.4.1	New User Link Prediction Problem Description	242
7.4.2	Cold Start Link Prediction Problem Formulation	244
7.4.3	Link Prediction Within Target Network	245
7.4.4	Cold-Start Link Prediction	250

7.5	Spy Technique Based Inter-Network PU Link Prediction	251
7.5.1	Cross-Network Concurrent Link Prediction Problem	251
7.5.2	Concurrent Link Prediction Problem Formulation	252
7.5.3	Social Meta Path Definition and Selection	253
7.5.4	Spy Technique Based PU Link Prediction	256
7.5.5	Multi-Network Concurrent PU Link Prediction Framework	258
7.6	Sparse and Low Rank Matrix Estimation Based PU Link Prediction	259
7.6.1	Problem Description	259
7.6.2	Intra-Network Link Prediction	261
7.6.3	Inter-Network Link Prediction	262
7.6.4	Proximal Operator Based CCCP Algorithm	267
7.7	Summary	269
7.8	Bibliography Notes	270
7.9	Exercises	271
	References	271
8	Community Detection	275
8.1	Overview	275
8.2	Traditional Homogeneous Network Community Detection	276
8.2.1	Node Proximity Based Community Detection	277
8.2.2	Modularity Maximization Based Community Detection	278
8.2.3	Spectral Clustering Based Community Detection	280
8.3	Emerging Network Community Detection	283
8.3.1	Background Knowledge	284
8.3.2	Problem Formulation	285
8.3.3	Intimacy Matrix of Homogeneous Network	286
8.3.4	Intimacy Matrix of Attributed Heterogeneous Network	287
8.3.5	Intimacy Matrix Across Aligned Heterogeneous Networks	290
8.3.6	Approximated Intimacy to Reduce Dimension	292
8.3.7	Clustering and Weight Self-adjustment	293
8.4	Mutual Community Detection	295
8.4.1	Background Knowledge	295
8.4.2	Problem Formulation	296
8.4.3	Meta Path Based Social Proximity Measure	297
8.4.4	Network Characteristic Preservation Clustering	298
8.4.5	Discrepancy Based Clustering of Multiple Networks	299
8.4.6	Joint Mutual Clustering of Multiple Networks	301
8.5	Large-Scale Network Synergistic Community Detection	304
8.5.1	Problem Formulation	304
8.5.2	Distributed Multilevel k-Way Partitioning	304
8.5.3	Distributed Synergistic Partitioning Process	306
8.6	Summary	308
8.7	Bibliography Notes	310
8.8	Exercises	311
	References	312

9	Information Diffusion	315
9.1	Overview	315
9.2	Traditional Information Diffusion Models	316
9.2.1	Threshold Based Diffusion Model	317
9.2.2	Cascade Based Diffusion Model	319
9.2.3	Epidemic Diffusion Model	321
9.2.4	Heat Diffusion Models	323
9.3	Intertwined Diffusion Models	324
9.3.1	Intertwined Diffusion Models for Multiple Topics	325
9.3.2	Diffusion Models for Signed Networks	328
9.4	Inter-Network Information Diffusion	331
9.4.1	Network Coupling Based Cross-Network Information Diffusion	332
9.4.2	Random Walk Based Cross-Network Information Diffusion	334
9.5	Information Diffusion Across Online and Offline World	337
9.5.1	Background Knowledge	337
9.5.2	Preliminary	339
9.5.3	Online Diffusion Channel	340
9.5.4	Offline Diffusion Channel	341
9.5.5	Hybrid Diffusion Channel	342
9.5.6	Channel Aggregation	343
9.5.7	Channel Weighting and Selection	344
9.6	Summary	345
9.7	Bibliography Notes	346
9.8	Exercises	347
	References	347
10	Viral Marketing	351
10.1	Overview	351
10.2	Traditional Influence Maximization	352
10.2.1	Influence Maximization Problem	352
10.2.2	Approximated Seed User Selection	353
10.2.3	Heuristics-Based Seed User Selection	356
10.3	Intertwined Influence Maximization	357
10.3.1	Conditional TIM	358
10.3.2	Joint TIM	360
10.4	Cross-Network Influence Maximization	365
10.4.1	Greedy Seed User Selection Across Networks	366
10.4.2	Dynamic Programming-Based Seed User Selection	369
10.5	Rumor Initiator Detection	373
10.5.1	The ISOMIT Problem	375
10.5.2	NP-Hardness of Exact ISOMIT Problem	375
10.5.3	A Special Case: k-ISOMIT-BT Problem	376
10.5.4	RID Method for General Networks	377
10.6	Summary	380
10.7	Bibliography Notes	381
10.8	Exercises	382
	References	382

11 Network Embedding	385
11.1 Overview	385
11.2 Relation Translation Based Graph Entity Embedding	386
11.2.1 TransE	386
11.2.2 TransH	387
11.2.3 TransR	389
11.3 Homogeneous Network Embedding	390
11.3.1 DeepWalk	390
11.3.2 LINE	393
11.3.3 node2vec	395
11.4 Heterogeneous Network Embedding	397
11.4.1 HNE: Heterogeneous Information Network Embedding	398
11.4.2 Path-Augmented Heterogeneous Network Embedding	400
11.4.3 HEBE: HyperEdge Based Embedding	400
11.5 Emerging Network Embedding Across Networks	403
11.5.1 Concept Definition and Problem Formulation	404
11.5.2 Deep DIME for Emerging Network Embedding	405
11.6 Summary	410
11.7 Bibliography Notes	411
11.8 Exercises	412
References	412

Part IV Future Directions

12 Frontier and Future Directions	417
12.1 Overview	417
12.2 Large-Scale Broad Learning	417
12.3 Multi-Source Broad Learning	418
12.4 Broad Learning Applications	418
12.5 Summary	418
12.6 Exercises	418
References	419