

Contents

1	Introduction	1
1.1	The Knowledge Discovery Process	1
1.2	Preprocessing	5
1.2.1	Data Preparation	5
1.2.2	Data Reduction	6
1.3	Data Mining	7
1.3.1	Supervised Learning	8
1.3.2	Unsupervised Learning	9
1.3.3	Semi-supervised Learning	10
1.3.4	Scalability Consideration	10
1.4	Classification	11
1.4.1	Validation Schemes	12
1.4.2	Evaluation Measures	14
	References	15
2	Multiple Instance Learning	17
2.1	Formal Description	17
2.2	Origin of MIL	18
2.2.1	Relationship with Propositional Learning	19
2.2.2	Relationship with Relational Learning	20
2.3	MIL Paradigms	20
2.3.1	Multi-instance Classification and Regression	21
2.3.2	Multi-instance Clustering	21
2.3.3	Instance Annotation	22
2.4	Applications of MIL	23
2.4.1	Bioinformatics	24
2.4.2	Image Classification and Retrieval	24
2.4.3	Web Mining and Text Classification	25
2.4.4	Object Detection and Tracking	25
2.4.5	Medical Diagnosis and Imaging	26

2.4.6	Other Classification Applications.	26
2.4.7	Regression Applications	27
2.4.8	Clustering Applications	28
	References.	29
3	Multi-instance Classification	35
3.1	Introduction	35
3.2	Formal Description	37
3.3	Taxonomy.	39
3.4	MI Assumptions	42
3.4.1	Standard MI Assumption.	44
3.4.2	Weidmann et al.'s Hierarchy	44
3.4.3	Collective Assumption.	45
3.4.4	Mixture Distribution Assumption	47
3.4.5	Soft Bag MI Assumption.	49
3.5	Distance Metrics	50
3.5.1	Bags as Point Sets.	51
3.5.2	Bags as Probability Distributions.	54
3.6	Real-World Applications.	56
3.6.1	Bioinformatics.	56
3.6.2	Image Classification and Retrieval.	57
3.6.3	Web Mining and Text Classification	59
3.6.4	Medical Diagnosis and Imaging	61
3.6.5	Acoustic Classification	62
3.7	Some Comments on Software Tools.	62
	References.	64
4	Instance-Based Classification Methods	67
4.1	Introduction	67
4.2	Wrapper Methods to Single-Instance Learning Algorithms.	68
4.3	Maximum Likelihood-Based Methods	70
4.3.1	Maximum Likelihood Principle	70
4.3.2	Diverse Density	71
4.3.3	Logistic Regression	73
4.3.4	Boosting	74
4.4	Decision Rules and Tree-Based Methods	75
4.5	Instance-Level SVM.	77
4.6	Neural Network-Based Methods	80
4.6.1	Feedforward Neural Networks.	80
4.6.2	Recurrent Neural Networks	82
4.6.3	Decision-Based Neural Networks	82
4.6.4	Network Combinations	82
4.7	Evolutionary Based Methods	83

- 4.8 Experimental Analysis 86
 - 4.8.1 Setup 86
 - 4.8.2 Results and Discussion 87
- 4.9 Summarizing Comments 93
- References 94
- 5 Bag-Based Classification Methods 99**
 - 5.1 Introduction 99
 - 5.2 Original Bag Space Methods 100
 - 5.2.1 Nearest Neighbor Methods 100
 - 5.2.2 Bag-Level SVM 102
 - 5.3 Mapped Bag Space Methods 103
 - 5.3.1 Mapping Methods Based on Bag Statistics 104
 - 5.3.2 Mapping Methods Based on Prototype Concatenation 106
 - 5.3.3 Mapping Methods Based on Counting 106
 - 5.3.4 Mapping Methods Based on Distance 112
 - 5.3.5 Bag-Level Distance Mapping Methods 115
 - 5.4 Experimental Analysis 115
 - 5.4.1 Setup 116
 - 5.4.2 Results and Discussion 117
 - 5.5 Comparing Instance-Based, Bag-Based, and Traditional Classification Methods 122
 - 5.6 Summarizing Comments 123
 - References 124
- 6 Multi-instance Regression 127**
 - 6.1 Introduction 127
 - 6.2 MIR Formulation 128
 - 6.2.1 Problem Description 128
 - 6.2.2 Evaluation Measures 128
 - 6.3 Instance-Based Regression Methods 129
 - 6.3.1 Prime Instance Assumption 130
 - 6.3.2 Collective Assumption 134
 - 6.4 Bag-Based Regression Methods 137
 - 6.4.1 Original Bag Space Methods 138
 - 6.4.2 Mapped Bag Space Methods 138
 - 6.5 Summarizing Comments 139
 - References 139
- 7 Unsupervised Multiple Instance Learning 141**
 - 7.1 Multiple Instance Cluster Analysis 141
 - 7.1.1 Introduction to Cluster Analysis 141
 - 7.1.2 Multiple Instance Clustering Requirements 145
 - 7.1.3 Multiple Instance Clustering Evaluation Measures 146

7.1.4	Multiple Instance Clustering Methods	148
7.1.5	Multiple Instance Clustering as a Preprocessing Step for Classification	159
7.2	Multiple Instance Association Rule Mining	160
7.2.1	Association Rule Mining Introduction	161
7.2.2	Multiple Instance Association Rule Mining Requirements	162
7.2.3	Apriori-MI Algorithm	164
7.3	Summarizing Comments	166
	References	166
8	Data Reduction	169
8.1	Introduction	169
8.2	Multiple Instance Methods for Feature Selection	170
8.2.1	Introduction to Feature Selection	171
8.2.2	Filter Methods	173
8.2.3	Embedded Methods	175
8.2.4	Hybrid Method: HyDR-MI Algorithm	181
8.3	Multiple Instance Methods for Bag Prototype Selection	182
8.3.1	Introduction to Bag Prototype Selection	182
8.3.2	Filter Methods	184
8.4	Summarizing Comments	187
	References	187
9	Imbalanced Multi-instance Data	191
9.1	Introduction	191
9.1.1	Dealing with Class Imbalance	192
9.1.2	Evaluation Measures in the Imbalanced Domain	193
9.2	Single-Instance SMOTE	194
9.3	Multi-instance Class Imbalance	194
9.3.1	Problem Description	195
9.3.2	Solutions for Multi-instance Class Imbalance	196
9.4	Multi-instance Resampling Methods	196
9.4.1	BagSMOTE, InstanceSMOTE, Bag_oversampling	196
9.4.2	B-Instances	198
9.4.3	B-Bags	199
9.5	Customized Multi-instance Approaches	199
9.5.1	Cost-Sensitive Boosting Models	200
9.5.2	Fuzzy Rough Multi-instance Classifiers	201
9.6	Experimental Analysis	201
9.6.1	Setup	202
9.6.2	Results and Discussion	202
9.7	Summarizing Comments	206
	References	206

- 10 Multiple Instance Multiple Label Learning 209**
 - 10.1 Introduction 209
 - 10.2 Formal Definition 211
 - 10.3 Applications 212
 - 10.3.1 Image Classification. 212
 - 10.3.2 Video and Audio Concept Detection. 213
 - 10.3.3 Text Categorization 214
 - 10.3.4 Bioinformatics. 214
 - 10.4 Evaluation Metrics 215
 - 10.5 Multi-instance Multi-label Learning Methods 216
 - 10.5.1 Methods Based on Problem Degeneration 217
 - 10.5.2 Methods Based on Problem Regularization 220
 - 10.6 Case Study: Kaggle Yelp Challenge. 223
 - 10.6.1 Dataset of Round 6 Yelp Challenge 224
 - 10.6.2 Winners of Round 6 Yelp Challenge. 225
 - 10.7 Relevant Multi-instance Multi-label Learning Research
Directions 226
 - 10.8 Summarizing Comments. 227
 - References. 227
- Glossary 231**



<http://www.springer.com/978-3-319-47758-9>

Multiple Instance Learning

Foundations and Algorithms

Herrera, F.; Ventura, S.; Bello, R.; Cornelis, C.; Zafra, A.;

Sánchez-Tarragó, D.; Vluymans, S.

2016, XI, 233 p. 46 illus., 40 illus. in color., Hardcover

ISBN: 978-3-319-47758-9