

# Contents

<b>1</b>	<b>Introduction</b>	1
1.1	Overview	1
1.2	The Knowledge Discovery in Databases Process	2
1.3	Data Preprocessing	3
1.4	Data Mining	6
1.4.1	DM Methods Attending to Available Data	6
1.4.2	DM Methods Attending to Target Objective	7
1.4.3	DM Methods Attending to Knowledge Representation	8
1.5	Classification	11
1.5.1	Binary Classification	11
1.5.2	Multiclass Classification	12
1.5.3	Multilabel Classification	13
1.5.4	Multidimensional Classification	14
1.5.5	Multiple Instance Learning	14
	References	15
<b>2</b>	<b>Multilabel Classification</b>	17
2.1	Introduction	17
2.2	Problem Formal Definition	18
2.2.1	Definitions	18
2.2.2	Symbols	18
2.2.3	Terminology	19
2.3	Applications of Multilabel Classification	19
2.3.1	Text Categorization	20
2.3.2	Labeling of Multimedia Resources	20
2.3.3	Genetics/Biology	21
2.3.4	Other Application Fields	21
2.3.5	MLDs Repositories	22

2.4	Learning from Multilabel Data . . . . .	22
2.4.1	The Data Transformation Approach. . . . .	23
2.4.2	The Method Adaptation Approach. . . . .	24
2.4.3	Ensembles of Classifiers. . . . .	25
2.4.4	Label Correlation Information. . . . .	26
2.4.5	High Dimensionality . . . . .	26
2.4.6	Label Imbalance . . . . .	27
2.5	Multilabel Data Tools . . . . .	28
	References . . . . .	29
<b>3</b>	<b>Case Studies and Metrics . . . . .</b>	<b>33</b>
3.1	Overview . . . . .	33
3.2	Case Studies . . . . .	34
3.2.1	Text Categorization . . . . .	34
3.2.2	Labeling of Multimedia Resources . . . . .	38
3.2.3	Genetics/Biology. . . . .	40
3.2.4	Synthetic MLDs . . . . .	41
3.3	MLD Characteristics . . . . .	41
3.3.1	Basic Metrics . . . . .	42
3.3.2	Imbalance Metrics. . . . .	43
3.3.3	Other Metrics . . . . .	44
3.3.4	Summary of Characterization Metrics . . . . .	45
3.4	Multilabel Classification by Example . . . . .	50
3.4.1	The ML-kNN Algorithm . . . . .	50
3.4.2	Experimental Configuration and Results . . . . .	51
3.5	Assessing Classifiers Performance . . . . .	54
3.5.1	Example-Based Metrics . . . . .	55
3.5.2	Label-based Metrics . . . . .	59
	References . . . . .	61
<b>4</b>	<b>Transformation-Based Classifiers . . . . .</b>	<b>65</b>
4.1	Introduction . . . . .	65
4.2	Multilabel Data Transformation Approaches . . . . .	66
4.3	Binary Classification Based Methods . . . . .	67
4.3.1	OVO Versus OVA Approaches. . . . .	67
4.3.2	Ensembles of Binary Classifiers . . . . .	68
4.4	Multiclass Classification-Based Methods. . . . .	69
4.4.1	Labelsets and Pruned Labelsets . . . . .	70
4.4.2	Ensembles of Multiclass Classifiers . . . . .	71
4.5	Data Transformation Methods in Practice . . . . .	72
4.5.1	Experimental Configuration . . . . .	72
4.5.2	Classification Results. . . . .	73
4.6	Summarizing Comments. . . . .	77
	References . . . . .	78

<b>5</b>	<b>Adaptation-Based Classifiers</b>	81
5.1	Overview	81
5.2	Tree-Based Methods	82
5.2.1	Multilabel C4.5, ML-C4.5	82
5.2.2	Multilabel Alternate Decision Trees, ADTBoost.MH	82
5.2.3	Other Tree-Based Proposals	83
5.3	Neuronal Network-Based Methods	83
5.3.1	Multilabel Back-Propagation, BP-MLL	83
5.3.2	Multilabel Radial Basis Function Network, ML-RBF	84
5.3.3	Canonical Correlation Analysis and Extreme Learning Machine, CCA-ELM	85
5.4	Vector Support Machine-Based Methods	85
5.4.1	MODEL-x	85
5.4.2	Multilabel SVMs Based on Ranking, Rank-SVM and SCRANK-SVM	86
5.5	Instance-Based Methods	86
5.5.1	Multilabel kNN, ML-kNN	86
5.5.2	Instance-Based and Logistic Regression, IBLR-ML	87
5.5.3	Other Instance-Based Classifiers	87
5.6	Probabilistic Methods	88
5.6.1	Collectible Multilabel Classifiers, CML and CMLF	88
5.6.2	Probabilistic Generic Models, PMM1 and PMM2	88
5.6.3	Probabilistic Classifier Chains, PCC	89
5.6.4	Bayesian and Tree Naïve Bayes Classifier Chains, BCC and TNBCC	89
5.6.5	Conditional Restricted Boltzmann Machines, CRBM	89
5.7	Other MLC Adaptation-Based Methods	90
5.8	Adapted Methods in Practice	91
5.8.1	Experimental Configuration	92
5.8.2	Classification Results	92
5.9	Summarizing Comments	97
	References	98
<b>6</b>	<b>Ensemble-Based Classifiers</b>	101
6.1	Introduction	101
6.2	Ensembles of Binary Classifiers	102
6.2.1	Ensemble of Classifier Chains, ECC	102
6.2.2	Ranking by Pairwise Comparison, RPC	102
6.2.3	Calibrated Label Ranking, CLR	103
6.3	Ensembles of Multiclass Classifiers	103
6.3.1	Ensemble of Pruned Sets, EPS	103
6.3.2	Random k-Labelsets, RAKEL	104
6.3.3	Hierarchy of Multilabel Classifiers, HOMER	104
6.4	Other Ensembles	104

6.5	Ensemble Methods in Practice. . . . .	105
6.5.1	Experimental Configuration . . . . .	106
6.5.2	Classification Results. . . . .	107
6.5.3	Training and Testing Times . . . . .	110
6.6	Summarizing Comments. . . . .	111
	References . . . . .	112
<b>7</b>	<b>Dimensionality Reduction. . . . .</b>	<b>115</b>
7.1	Overview . . . . .	115
7.1.1	High-Dimensional Input Space . . . . .	116
7.1.2	High-Dimensional Output Space . . . . .	117
7.2	Feature Space Reduction . . . . .	117
7.2.1	Feature Engineering Approaches . . . . .	118
7.2.2	Multilabel Supervised Feature Selection . . . . .	119
7.2.3	Experimentation . . . . .	121
7.3	Label Space Reduction. . . . .	124
7.3.1	Sparseness and Dependencies Among Labels . . . . .	124
7.3.2	Proposals for Reducing Label Space Dimensionality . . . . .	125
7.3.3	Experimentation . . . . .	126
7.4	Summarizing Comments. . . . .	129
	References . . . . .	129
<b>8</b>	<b>Imbalance in Multilabel Datasets . . . . .</b>	<b>133</b>
8.1	Introduction . . . . .	133
8.2	Imbalanced MLD Specificities. . . . .	134
8.2.1	How to Measure the Imbalance Level . . . . .	135
8.2.2	Concurrence Among Imbalanced Labels. . . . .	136
8.3	Facing Imbalanced Multilabel Classification . . . . .	138
8.3.1	Classifier Adaptation . . . . .	138
8.3.2	Resampling Techniques . . . . .	139
8.3.3	The Ensemble Approach . . . . .	145
8.4	Multilabel Imbalanced Learning in Practice. . . . .	146
8.4.1	Experimental Configuration . . . . .	147
8.4.2	Classification Results. . . . .	147
8.5	Summarizing Comments. . . . .	150
	References . . . . .	150
<b>9</b>	<b>Multilabel Software . . . . .</b>	<b>153</b>
9.1	Overview . . . . .	153
9.2	Working with Multilabel Data. . . . .	154
9.2.1	Multilabel Data File Formats . . . . .	154
9.2.2	Multilabel Data Repositories. . . . .	155
9.2.3	The mldr.datasets Package . . . . .	157
9.2.4	Generating Synthetic MLDs . . . . .	162

9.3	Exploratory Analysis of MLDs . . . . .	162
9.3.1	MEKA. . . . .	163
9.3.2	The mldr Package . . . . .	166
9.4	Conducting Multilabel Experiments . . . . .	179
9.4.1	MEKA. . . . .	179
9.4.2	MULAN . . . . .	182
9.4.3	The RunMLClassifier Utility. . . . .	188
9.5	Summarizing Comments. . . . .	189
	References . . . . .	190
	<b>Glossary . . . . .</b>	<b>193</b>

Multilabel Classification

Problem Analysis, Metrics and Techniques

Herrera, F.; Charte, F.; Rivera, A.J.; del Jesus, M.J.

2016, XVI, 194 p. 72 illus., Hardcover

ISBN: 978-3-319-41110-1