Contents

Preface vii

Acknowledgments ix

1. Introduction and Basic Concepts 1
   - 1.1 Pattern Recognition 1
   - 1.2 Learning Methodology 3
   - 1.3 Statistical and Structural Pattern Recognition 6
   - 1.4 Dissimilarity Representation for Pattern Recognition . . . 9
   - 1.5 Summary and Outline 12

2. Graph Matching 15
   - 2.1 Graph and Subgraph 17
   - 2.2 Exact Graph Matching 19
   - 2.3 Error-tolerant Graph Matching 27
   - 2.4 Summary and Broader Perspective 32

3. Graph Edit Distance 35
   - 3.1 Basic Definition and Properties 37
     - 3.1.1 Conditions on Edit Cost Functions 39
     - 3.1.2 Examples of Edit Cost Functions 41
   - 3.2 Exact Computation of GED 44
   - 3.3 Efficient Approximation Algorithms 46
     - 3.3.1 Bipartite Graph Matching 47
     - 3.3.2 Graph Edit Distance Computation by Means of Munkres' Algorithm 52

GRAPH CLASSIFICATION AND CLUSTERING BASED ON VECTOR SPACE EMBEDDING
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/compsci/7731.html

3.4 Exact vs. Approximate Graph Edit Distance – An Experimental Evaluation . . . 55
3.4.1 Nearest-Neighbor Classification . . . 55
3.4.2 Graph Data Set . . . 56
3.4.3 Experimental Setup and Validation of the Meta Parameters . . . 57
3.4.4 Results and Discussion . . . 58
3.5 Summary . . . 63

4. Graph Data 65

4.1 Graph Data Sets . . . 66
4.1.1 Letter Graphs . . . 66
4.1.2 Digit Graphs . . . 68
4.1.3 GREC Graphs . . . 71
4.1.4 Fingerprint Graphs . . . 74
4.1.5 AIDS Graphs . . . 77
4.1.6 Mutagenicity Graphs . . . 78
4.1.7 Protein Graphs . . . 79
4.1.8 Webpage Graphs . . . 81
4.2 Evaluation of Graph Edit Distance . . . 83
4.3 Data Visualization . . . 92
4.4 Summary . . . 95

5. Kernel Methods 97

5.1 Overview and Primer on Kernel Theory . . . 97
5.2 Kernel Functions . . . 98
5.3 Feature Map vs. Kernel Trick . . . 104
5.4 Kernel Machines . . . 110
5.4.1 Support Vector Machine (SVM) . . . 110
5.4.2 Principal Component Analysis (PCA) . . . 117
5.4.3 k-Means Clustering . . . 122
5.5 Graph Kernels . . . 125
5.6 Experimental Evaluation . . . 129
5.7 Summary . . . 131

6. Graph Embedding Using Dissimilarities 133

6.1 Related Work . . . 135
6.1.1 Graph Embedding Techniques . . . 135

GRAPH CLASSIFICATION AND CLUSTERING BASED ON VECTOR SPACE EMBEDDING
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/compsci/7731.html

6.1.2 Dissimilarities as a Representation Formalism . . 137
6.2 Graph Embedding Using Dissimilarities 139
6.2.1 General Embedding Procedure and Properties . . 139
6.2.2 Relation to Kernel Methods 142
6.2.3 Relation to Lipschitz Embeddings 144
6.2.4 The Problem of Prototype Selection 146
6.3 Prototype Selection Strategies 148
6.4 Prototype Reduction Schemes 157
6.5 Feature Selection Algorithms 163
6.6 Defining the Reference Sets for Lipschitz Embeddings . . 170
6.7 Ensemble Methods 171
6.8 Summary 173

7. Classification Experiments with Vector Space Embedded Graphs 175
7.1 Nearest-Neighbor Classifiers Applied to Vector Space Embedded Graphs 176
7.2 Support Vector Machines Applied to Vector Space Embedded Graphs 181
7.2.1 Prototype Selection 181
7.2.2 Prototype Reduction Schemes 192
7.2.3 Feature Selection and Dimensionality Reduction . 195
7.2.4 Lipschitz Embeddings 205
7.2.5 Ensemble Methods 210
7.3 Summary and Discussion 214

8. Clustering Experiments with Vector Space Embedded Graphs 221
8.1 Experimental Setup and Validation of the Meta parameters 222
8.2 Results and Discussion 226
8.3 Summary and Discussion 231

9. Conclusions 235

Appendix A Validation of Cost Parameters 247

Appendix B Visualization of Graph Data 255

Appendix C Classifier Combination 259

GRAPH CLASSIFICATION AND CLUSTERING BASED ON VECTOR SPACE EMBEDDING
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/compsci/7731.html

Appendix D Validation of a k-NN classifier in the Embedding Space 263

Appendix E Validation of a SVM classifier in the Embedding Space 273

Appendix F Validation of Lipschitz Embeddings 277

Appendix G Validation of Feature Selection Algorithms and PCA Reduction 289

Appendix H Validation of Classifier Ensemble 293

Appendix I Validation of Kernel k-Means Clustering 295

Appendix J Confusion Matrices 305

Bibliography 309

Index 329

GRAPH CLASSIFICATION AND CLUSTERING BASED ON VECTOR SPACE EMBEDDING
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/compsci/7731.html