

Contents

Foreword	vii
Preface	xi
1. Introduction to Bioinformatics	1
1.1 What Is Bioinformatics	1
1.2 A Brief History of Bioinformatics	2
1.3 Scope of Bioinformatics	8
1.4 Major Challenges in Bioinformatics	13
1.5 Bioinformatics and Computer Science	14
2. Introduction to Fuzzy Set Theory and Fuzzy Logic	16
2.1 Where Does Fuzzy Logic Fit in Computational Science?	16
2.2 Why Do We Need to Use Fuzziness in Biology?	17
2.3 Brief History of the Field	21
2.4 Fuzzy Membership Functions and Operators	23
2.4.1 Membership functions	23
2.4.2 Basic fuzzy set operators	27
2.4.3 Compensatory operators	33
2.5 Fuzzy Relations and Fuzzy Logic Inference	37
2.6 Fuzzy Clustering	48
2.6.1 Fuzzy C-Means	49
2.6.2 Extension to fuzzy C-Means	54
2.6.3 Possibilistic C-Means (PCM)	59
2.7 Fuzzy K-Nearest Neighbors	63
2.8 Fuzzy Measures and Fuzzy Integrals	66
2.8.1 Fuzzy measures	67
2.8.2 Fuzzy integrals	69
2.9 Summary and Final Thoughts	72

3. Fuzzy Similarities in Ontologies	73
3.1 Introduction	73
3.2 Definition of Ontology-Based Similarity	76
3.3 Set-Based Similarity Measure	80
3.3.1 Pair-wise aggregation	80
3.3.2 Bag of words similarities	84
3.4 Fuzzy Measure Similarity	85
3.5 Fuzzy Measure Similarity for Augmented Sets of Ontology Objects	86
3.6 Choquet Fuzzy Integral Similarity Measure	87
3.7 Examples and Applications of Fuzzy Measure Similarity Using GO Terms	90
3.7.1 Lymphoma case study	90
3.7.2 Gene clustering using Gene Ontology annotations	92
3.7.3 Gene summarization using Gene Ontology terms	98
3.8 Ontology Similarity in Data Mining	99
3.9 Discussion and Summary	102
4. Fuzzy Logic in Structural Bioinformatics	103
4.1 Introduction	103
4.2 Protein Secondary Structure Prediction	106
4.3 Protein Solvent Accessibility Prediction	116
4.4 Protein Structure Matching Using Fuzzy Alignments	118
4.5 Protein Similarity Calculation Using Fuzzy Contact Maps	124
4.6 Protein Structure Class Classification	126
4.7 Summary	130
5. Application of Fuzzy Logic in Microarray Data Analyses	131
5.1 Introduction	131
5.1.1 Microarray data description	134
5.1.2 Microarray processing algorithms for gene selection and patient classification	136
5.1.3 Microarray processing algorithms for gene regulatory network discovery	137
5.2 Clustering Algorithms	138
5.2.1 (Dis)similarity measures for microarray data	140
5.2.2 Fuzzy C-means (FCM)	144
5.2.3 Relational fuzzy C-means	148
5.2.4 Fuzzy co-clustering algorithms	152
5.3 Inferring Gene Networks Using Fuzzy Rule Systems	155
5.4 Discussion and Summary	159

6. Other Applications	160
6.1 Overview	160
6.2 Applications in Biological Sequence Analyses	161
6.2.1 Protein sequence comparison	161
6.2.2 Application in sequence family classification	164
6.2.3 Application in motif identification	165
6.2.4 Application in protein subcellular localization prediction	166
6.2.5 Genomic structure prediction	167
6.3 Application in Computational Proteomics	168
6.3.1 Electrophoresis analysis	169
6.3.2 Protein identification through mass-spec	170
6.4 Application in Drug Design	171
6.5 Discussion and Summary	174
7. Summary and Outlook	176
Appendix I. Fundamental Biological Concepts	179
AI.1 DNA, RNA and Genome	179
AI.1.1 DNA (deoxyribonucleic acid)	179
AI.1.2 RNA (ribonucleic acid)	181
AI.1.3 Genome	181
AI.2 Protein and Its Structure	182
AI.3 Central Dogma of Biology	186
Appendix II. Online Resources	189
AII.1 Online Resources for Molecular Biology	189
AII.2 Online Resources for Bioinformatics	190
AII.2.1 Protein structure	191
AII.2.1.1 Protein structure database	191
AII.2.1.2 Protein structure visualization	191
AII.2.2 Microarray	192
AII.2.2.1 Microarray databases	192
AII.2.2.2 Microarray analysis tool	193
AII.2.3 Gene ontology	194
AII.2.4 Online portals for bioinformatics	194
AII.3 Online Resources for Fuzzy Set Theory and Fuzzy Logic	195
Bibliography	196
Index	222