
Contents

1 Computational Intelligence Algorithms and DNA Microarrays

<i>D.K. Tasoulis, V.P. Plagianakos, and M.N. Vrahatis</i>	1
1.1 Introduction	1
1.2 Neural Networks	2
1.2.1 The Rprop Neural Network Training Algorithm	3
1.2.2 The Adaptive Online Neural Network Training Algorithm	4
1.3 Evolutionary Algorithms	5
The Differential Evolution Algorithm	5
1.4 Feature Selection and Dimension Reduction Techniques	7
1.4.1 Principal Component Analysis	7
1.4.2 Reducing the Dimensions using Clustering	7
Unsupervised k -Windows Clustering Algorithm	9
The DBSCAN Clustering Algorithm	10
The Fuzzy c -Means Clustering Algorithm	12
1.4.3 The PDDP Clustering Algorithm	12
1.4.4 Growing Neural Gas	13
1.4.5 A Hybrid Approach	13
1.5 Experimental Analysis	14
1.5.1 DNA Microarray Problems	14
1.5.2 Evaluation of the Clustering Algorithms	15
Clustering Based on Supervised Gene Selection	15
Clustering Based on Unsupervised Gene Selection	18
1.5.3 Using Clustering and FNN classifiers	21
1.5.4 Using Evolutionary Algorithms and FNN classifiers	24
1.6 Concluding Remarks	26
References	27

2 Inferring Gene Regulatory Networks from Expression Data

<i>Lars Kaderali and Nicole Radde</i>	33
2.1 Introduction	33
2.1.1 Biological Background	34
2.1.2 Modeling Gene Regulatory Networks	37

2.2	Boolean Networks	38
2.2.1	Inferring a Boolean Network from Data	38
	The REVerse Engineering ALgorithm REVEAL	40
2.2.2	Advantages and Disadvantages of the Boolean Network Model	41
2.2.3	Extensions of the Boolean Model	43
2.3	Relevance Networks and Information Theoretic Approaches	44
2.3.1	The ARACNe Algorithm	44
2.3.2	Advantages and Disadvantages of Relevance Network Approaches	44
2.4	Bayesian Networks	45
2.4.1	Learning a Bayesian Network from Data	47
	Learning Probabilities for a given Network Topology	48
	Learning the Network Topology	48
2.4.2	Advantages and Disadvantages of the Bayesian Network Model	49
2.4.3	Extensions of Bayesian Networks	50
2.5	Quantitative Models using Ordinary Differential Equations	50
	Linear Models	51
	Additive Models based on Chemical Reaction Kinetics	52
	S-Systems	55
2.5.1	Network Inference	56
2.5.2	Advantages and Disadvantages of ODE Models	57
2.6	Bayes Regularized Ordinary Differential Equations	58
2.6.1	Prior Distributions for Network Parameters	60
2.6.2	Application to the Yeast Cell Cycle Network	62
	Results	63
2.7	Other Approaches	64
2.7.1	Delay Differential Equations	64
2.7.2	Partial Differential Equations	66
2.7.3	Stochastic Kinetic Approaches	66
2.8	Conclusion	67
	References	69
3 Belief Networks for Bioinformatics		
	<i>Jeroen (H.H.L.M.) Donkers and Karl Tuyls</i>	75
3.1	An Introduction to Belief Networks	76
3.1.1	Belief Networks Defined	76
3.1.2	Belief Network Example	77
3.1.3	D-Separation	78
3.1.4	Local Probabilities	79
3.1.5	Markov Random Fields	80
3.2	Basic Algorithms for Belief Networks	80
3.2.1	Querying Belief Networks	81
3.2.2	Exact Inference by Bucket Elimination	81
3.2.3	Exact Inference in Polytrees by Belief Propagation	82

3.2.4 Exact Inference by Clique-Tree Belief Propagation	84
3.2.5 Exact Inference by Arithmetic Circuits	85
3.2.6 Approximate Inference by Sampling	86
3.2.7 Parameter Learning	87
3.2.8 Structure Learning	88
3.3 Variants and Extensions	89
3.3.1 Dynamic Belief Networks	89
3.3.2 Hidden Markov Models	90
3.3.3 Qualitative Belief Networks	92
3.3.4 Probabilistic Relational Models and Relational Belief Networks	93
3.4 Building Belief Networks by Hand	94
3.5 Applications in Bioinformatics	96
3.5.1 Sequence Analysis, Protein Prediction, and Motif Finding	96
3.5.2 Gene Regulation Networks	98
3.5.3 Other Cellular Networks	100
3.5.4 Linkage Analysis and Genetics	102
3.5.5 Discussion	103
3.6 Software Tools	103
3.6.1 Integrated Tools and Suites	104
3.6.2 Partial Tools	105
3.6.3 APIs, Libraries, and Toolkits	105
3.6.4 Belief Network Tools for Bioinformatics	106
3.7 Guide to the Literature	106
3.8 Conclusion	107
References	107
 4 Swarm Intelligence Algorithms in Bioinformatics	
<i>Swagatam Das, Ajith Abraham, and Amit Konar</i>	113
4.1 Introduction	113
4.2 Fundamental Concepts in Bioinformatics	115
4.2.1 DNA	115
4.2.2 The Gene	115
4.2.3 Proteins	116
4.2.4 DNA Microarray	118
4.3 Swarm Intelligence - an Overview and Relevance to Bioinformatics	119
4.3.1 Ant Colony Systems	123
4.3.2 The ACO Algorithm	124
4.3.3 The Particle Swarm Optimisation (PSO)	126
4.3.4 Relevance of SI Algorithms in Bioinformatics	129
4.4 A Review of the Present State of the Art	130
4.4.1 Clustering of Gene Expression Data	130
4.4.2 The Molecular Docking Problem	131
4.4.3 The Multiple Sequence Alignment Problems (MSA)	133

4.4.4 The Construction of Phylogenetic Trees	135
4.4.5 The RNA Secondary Structure Prediction	137
4.4.6 Protein Secondary Structure Prediction	138
4.4.7 Fragment Assembly Problem (FAP)	139
4.5 A Few Open-ended Problems and Future Research Directions	140
4.5.1 Identifying Gene Regulatory Networks	140
4.5.2 Protein Tertiary Structure Prediction and Folding	140
4.5.3 Characterization of Metabolic Pathways between Different Genomes	141
4.5.4 Characterization of Metabolic Pathways between Different Genomes	141
4.6 Conclusions	142
References	143
 5 Time Course Gene Expression Classification with Time Lagged Recurrent Neural Network	
<i>Yulan Liang and Arpad Kelemen</i>	149
5.1 Introduction	149
5.2 Data Acquisition and Preprocessing	152
5.2.1 Data Extraction	152
5.2.2 Functional Class Extraction	152
5.2.3 Data Preprocessing	153
5.2.4 Smoothing the Data	154
5.2.5 Input Selection	154
5.3 Design of Time Lagged Recurrent Neural Network	156
5.3.1 Memory Structures	156
5.3.2 Learning Algorithms	157
5.3.3 Statistical Criteria for Model Selection	158
5.4 Experimental Results	158
5.4.1 Two Classes: Ribosome and Non-Ribosome Protein Genes	158
5.4.2 Three Classes: Ribosomal, Transcription and Secretion Gene Functional Classes	158
5.4.3 Multiple Functional Classes	160
5.5 Related Works	160
5.6 Conclusion	161
References	162
 6 Tree-Based Algorithms for Protein Classification	
<i>Róbert Busa-Fekete, András Kocsor, and Sándor Pongor</i>	165
6.1 Introduction	166
6.2 Related Work	167
6.3 <i>TreeNN</i> : Protein Classification via Neighborhood Evaluation within Weighted Binary Trees	168
6.3.1 Conceptual Outline	168
6.3.2 Description of the Algorithm	170

6.3.3 Implementation	171
6.3.4 Performance Evaluation	171
6.4 <i>TreeInsert</i> : Protein Classification via Insertion into Weighted Binary Trees	174
6.4.1 Conceptual Outline	174
6.4.2 Description of the Algorithm	176
6.4.3 Implementation	177
6.4.4 Performance Evaluation	177
6.5 Discussion and Conclusions	179
6.6 Appendix: Datasets and Methods	179
6.6.1 Datasets and Classification Tasks	179
6.6.2 Sequence Comparison Algorithms	180
6.6.3 Distance-Based Tree Building Methods	180
6.6.4 Evaluation of Classification Performance (ROC, AUC and Error rate)	181
6.7 Acknowledgements	181
References	182
7 Covariance-Model-Based RNA Gene Finding: Using Dynamic Programming versus Evolutionary Computing	
<i>Scott F. Smith</i>	183
7.1 Introduction	183
7.2 Review of Covariance Models for RNA Gene Finding	184
7.2.1 CM Model Structure Determination	185
7.2.2 Mapping a Database Sequence to a Covariance Model	191
7.3 Application of CM-Based RNA Gene Finding: The Rfam Database	195
7.4 Filters to Reduce Database Search Cost	196
7.5 An Alternative to Filters Using Evolutionary Computation	197
7.5.1 Dynamic Programming Efficiency	197
7.5.2 CM-Based Search Without Dynamic Programming	199
7.5.3 Experimental Results	203
7.6 Conclusions and Future Direction	206
References	207
8 Fuzzy Classification for Gene Expression Data Analysis	
<i>Gerald Schaefer, Tomoharu Nakashima, and Yasuyuki Yokota</i>	209
8.1 Introduction	209
8.2 Methods	210
8.2.1 Fuzzy Rule Generation	211
8.2.2 Fuzzy Reasoning	212
8.2.3 Rule splitting	212
8.3 Results and Discussion	213
8.4 Conclusions	217
References	217

9 Towards the Enhancement of Gene Selection Performance

<i>D. Huang and Tommy W. S. Chow</i>	219
9.1 Introduction	219
9.2 Brief Reviews on Gene Selection	220
9.3 Modifying Search Engines	223
9.4 Point Injection for Gene Selection	225
9.5 Implementation	227
9.6 Experimental Results	227
9.6.1 Synthetic Data	228
9.6.2 Patience Data	228
Colon tumor classification	229
Prostate cancer classification	229
Leukemia subtype classification	230
Analysis on the stability of MSFS and WMSFS	230
Comparisons of SFS, MSFS and WMSFS	230
Detailed results on prostate cancer	233
9.7 Conclusions	233
9.8 Acknowledgement	235
References	235

10 *Saccharomyces pombe* and *Saccharomyces cerevisiae***Gene Regulatory Network Inference Using the Fuzzy****Logic Network**

<i>Yingjun Cao, Paul P. Wang, and Alade Tokuta</i>	237
10.1 Introduction	237
10.2 Fuzzy Logic Network Theory	238
10.3 Algorithm	243
10.4 Applications to Yeast Datasets	244
10.4.1 Data Pre-processing	245
10.4.2 Inference Results of the <i>S. cerevisiae</i> Dataset	246
10.4.3 Inference Results of the <i>S. pombe</i> Dataset	246
10.5 Conclusions and Discussion	250
10.6 Acknowledgment	254
References	254

11 Multivariate Regression Applied to Gene**Expression Dynamics**

<i>Olli Haavisto and Heikki Hyötyniemi</i>	257
11.1 Introduction	257
11.2 Approach	258
11.2.1 Linearity	258
11.2.2 Degrees of Freedom	259
11.3 Multivariate Regression Tools	259
11.3.1 Principal Component Analysis	260

11.3.2 Subspace Identification	261
Stochastic SSI Algorithm	262
Combined Deterministic-Stochastic SSI Algorithm	263
Extensions and Modifications	264
11.4 Examples	265
11.4.1 Data	265
11.4.2 Results	267
Stress Response	267
Cell Cycle	270
11.5 Conclusions	272
11.6 Acknowledgments	274
References	274

12 The Amine System Project: Systems Biology in Practice

<i>Ismael Navas-Delgado, Raúl Montañez, Miguel Ángel Medina, José Luis Urdiales, José F. Aldana, and Francisca Sánchez-Jiménez</i>	277
12.1 Introduction: Problem Statement and Development of a Systems Biology Pilot Project	277
12.2 The Semantic Web: Practical Usage in Systems Biology	280
12.3 Architecture and Future Prospects	282
12.3.1 The Pilot: AMine Metabolism Ontology and the Semantic Directory Core	282
12.3.2 Ontology-Based Mediation	284
12.3.3 Conclusion	287
12.4 Acknowledgements	289
References	290

13 DNA Encoding Methods in the Field of DNA Computing

<i>Aili Han and Daming Zhu</i>	293
13.1 Introduction	294
13.2 Preliminaries to DNA Computing	295
13.2.1 Orientation of DNA Molecule	295
13.2.2 Basic Operations in DNA Computing	296
13.3 DNA Encoding Methods for the Problems Related to Graphs without Weights	297
13.3.1 DNA Encoding Method for the Hamiltonian Path Problem	297
The Hamiltonian Path Problem	297
DNA Encoding Method	298
DNA Algorithm	298
13.3.2 DNA Encoding Method for the SAT Problem	298
The SAT Problem	299
DNA Encoding Method	299
DNA Algorithm	300

13.3.3 DNA Encoding Method for the Maximal Clique Problem	300
The Maximal Clique Problem	301
DNA Encoding Method	301
DNA Algorithm	301
13.4 DNA Encoding Methods for the Problems Related to Graphs	
with Weights	302
13.4.1 Encoding Weights by the Lengths of DNA Strands	303
DNA Encoding Method	303
DNA Algorithm	303
13.4.2 Encoding Weights by the Number of Hydrogen Bonds	303
DNA Encoding Method	304
DNA Algorithm	304
Algorithm for genetic code optimization	304
Molecular algorithm	305
13.4.3 Encoding Weights by the Concentrations of DNA Strands	305
DNA Encoding Method	305
DNA Algorithm	306
13.4.4 Encoding Weights by the Melting Temperatures	
of DNA Strands	306
DNA Encoding Method	306
DNA Algorithm	306
13.4.5 Encoding Weights by Means of the General Line Graph	307
The Chinese Postman Problem	307
Construction of General Line Graph	307
DNA Encoding Method	308
DNA Algorithm	311
13.4.6 RLM: Relative Length Method of Encoding Weights	311
Definitions Involved in the RLM Method	311
RLM Method of Encoding Weights	312
DNA Algorithm	313
13.4.7 Method of Encoding Nonlinear Solutions	314
The Minimum Spanning Tree Problem	314
DNA Encoding Method	314
DNA Algorithm	315
13.4.8 DNA Encoding Method for the Maximal Weight	
Clique Problem	317
DNA Encoding Method	317
DNA Algorithm	317
13.4.9 DNA Encoding Method for the 0/1 Knapsack Problem	318
DNA Encoding Method	318
DNA Algorithm	319
13.5 Conclusion	320
References	321
Index	323



<http://www.springer.com/978-3-540-76802-9>

Computational Intelligence in Bioinformatics

Kelemen, A.; Abraham, A.; Chen, Y. (Eds.)

2008, XVI, 326 p., Hardcover

ISBN: 978-3-540-76802-9