Pattern Recognition in Computational Molecular Biology. Techniques and Approaches. Wiley Series in Bioinformatics

Description: A comprehensive overview of high-performance pattern recognition techniques and approaches to Computational Molecular Biology

This book surveys the developments of techniques and approaches on pattern recognition related to Computational Molecular Biology. Providing a broad coverage of the field, the authors cover fundamental and technical information on these techniques and approaches, as well as discussing their related problems. The text consists of twenty nine chapters, organized into seven parts: Pattern Recognition in Sequences, Pattern Recognition in Secondary Structures, Pattern Recognition in Tertiary Structures, Pattern Recognition in Quaternary Structures, Pattern Recognition in Microarrays, Pattern Recognition in Phylogenetic Trees, and Pattern Recognition in Biological Networks.

- Surveys the development of techniques and approaches on pattern recognition in biomolecular data
- Discusses pattern recognition in primary, secondary, tertiary and quaternary structures, as well as microarrays, phylogenetic trees and biological networks
- Includes case studies and examples to further illustrate the concepts discussed in the book

Pattern Recognition in Computational Molecular Biology: Techniques and Approaches is a reference for practitioners and professional researchers in Computer Science, Life Science, and Mathematics. This book also serves as a supplementary reading for graduate students and young researchers interested in Computational Molecular Biology.

Mourad Elloumi, PhD, is Professor in Computer Science at the University of Tunis–El Manar, Tunisia. Dr. Elloumi is the author/co–author of more than 50 publications in international journals and conference proceedings related to Algorithmics, Computational Molecular Biology, and Knowledge Discovery and Data Mining.

Costas S. Iliopoulos, PhD, is Professor of Algorithm Design at King's College London, UK. Dr. Iliopoulos co–authored over 300 peer–reviewed articles in pattern matching and combinatorics of strings. He serves on the editorial board of the Journal of Discrete Algorithms, Computer Mathematics & Combinatorial Computing, and System Biology & Biomedical Technologies.

Jason T. L. Wang, PhD, is Professor of Computer Science at the New Jersey Institute of Technology, USA. Dr. Wang has published extensively on Data Mining and Computational Molecular Biology, and has been a member of program committees for over 200 conferences and workshops in these and related areas.

Albert Y. Zomaya, PhD, is the Chair Professor of High Performance Computing & Networking in the School of Information Technologies, University of Sydney, Australia. Dr. Zomaya published more than 500 scientific papers and articles and is author, co–author or editor of more than 20 books. Dr. Zomaya is Fellow of AAAS, IEEE, and IET.

Contents:

LIST OF CONTRIBUTORS xxi

PREFACE xxvii

I PATTERN RECOGNITION IN SEQUENCES 1

1 COMBINATORIAL HAPLOTYPING PROBLEMS 3
Giuseppe Lancia

1.1 Introduction / 3

1.2 Single Individual Haplotyping / 5
1.2.1 The Minimum Error Correction Model / 8
1.2.2 Probabilistic Approaches and Alternative Models / 10
1.3 Population Haplotyping / 12
1.3.1 Clark’s Rule / 14
1.3.2 Pure Parsimony / 15
1.3.3 Perfect Phylogeny / 19
1.3.4 Disease Association / 21
1.3.5 Other Models / 22

References / 23

2 ALGORITHMIC PERSPECTIVES OF THE STRING BARCODING PROBLEMS 28
Sima Behpour and Bhaskar DasGupta

2.1 Introduction / 28
2.2 Summary of Algorithmic Complexity Results for Barcoding Problems / 32
2.2.1 Average Length of Optimal Barcodes / 33
2.3 Entropy–Based Information Content Technique for Designing Approximation Algorithms for String Barcoding Problems / 34
2.4 Techniques for Proving Inapproximability Results for String Barcoding Problems / 36
2.4.1 Reductions from Set Covering Problem / 36
2.4.2 Reduction from Graph–Coloring Problem / 38
2.5 Heuristic Algorithms for String Barcoding Problems / 39
2.5.1 Entropy–Based Method with a Different Measure for Information Content / 39
2.5.2 Balanced Partitioning Approach / 40
2.6 Conclusion / 40

Acknowledgments / 41

References / 41

3 ALIGNMENT–FREE MEASURES FOR WHOLE–GENOME COMPARISON 43
Matteo Comin and Davide Verzotto

3.1 Introduction / 43
3.2 Whole–Genome Sequence Analysis / 44
3.2.1 Background on Whole–Genome Comparison / 44
3.2.2 Alignment–Free Methods / 45
3.2.3 Average Common Subword / 46
3.2.4 Kullback–Leibler Information Divergence / 47
7.3.1 Base Classifiers / 117
7.3.2 Ensemble Methods / 118
7.3.3 Convex Combination / 119
7.4 Postprocessing with A Generative Model / 119
7.5 Dedication to Protein Secondary Structure Prediction / 120
7.5.1 Biological Problem / 121
7.5.2 MSVMpred2 / 121
7.5.3 Hidden Semi–Markov Model / 122
7.5.4 Experimental Results / 122
7.6 Conclusions and Ongoing Research / 125
Acknowledgments / 126
References / 126

8 STRUCTURAL MOTIF IDENTIFICATION AND RETRIEVAL: A GEOMETRICAL APPROACH 129
Virginio Cantoni, Marco Ferretti, Mirto Musci, and Nahumi Nugrahaningsih
8.1 Introduction / 129
8.2 A Few Basic Concepts / 130
8.2.1 Hierarchy of Protein Structures / 130
8.2.2 Secondary Structure Elements / 131
8.2.3 Structural Motifs / 132
8.2.4 Available Sources for Protein Data / 134
8.3 State of the Art / 135
8.3.1 Protein Structure Motif Search / 135
8.3.2 Promotif / 136
8.3.3 Secondary–Structure Matching / 137
8.3.4 Multiple Structural Alignment by Secondary Structures / 138
8.4 A Novel Geometrical Approach to Motif Retrieval / 138
8.4.1 Secondary Structures Cooccurrences / 138
8.4.2 Cross Motif Search / 143
8.4.3 Complete Cross Motif Search / 146
8.5 Implementation Notes / 149
8.5.1 Optimizations / 149
8.5.2 Parallel Approaches / 150
16 STRUCTURAL DOMAINS IN PREDICTION OF BIOLOGICAL PROTEIN PROTEIN INTERACTIONS 291
Mina Maleki, Michael Hall, and Luis Rueda

16.1 Introduction / 291
16.2 Structural Domains / 293
16.3 The Prediction Framework / 293
16.4 Feature Extraction and Prediction Properties / 294
16.4.1 Physicochemical Properties / 296
16.4.2 Domain–Based Properties / 298
16.5 Feature Selection / 299
16.5.1 Filter Methods / 299
16.5.2 Wrapper Methods / 301
16.6 Classification / 301
16.6.1 Linear Dimensionality Reduction / 301
16.6.2 Support Vector Machines / 303
16.6.3 k–Nearest Neighbor / 303
16.6.4 Naive Bayes / 304
16.7 Evaluation and Analysis / 304
16.8 Results and Discussion / 304
16.8.1 Analysis of the Prediction Properties / 304
16.8.2 Analysis of Structural DDIs / 307
16.9 Conclusion / 309
References / 310

V PATTERN RECOGNITION IN MICROARRAYS 315

17 CONTENT–BASED RETRIEVAL OF MICROARRAY EXPERIMENTS 317
Hasan O¢gul

17.1 Introduction / 317
17.2 Information Retrieval: Terminology and Background / 318
17.3 Content-Based Retrieval / 320
17.4 Microarray Data and Databases / 322
17.5 Methods for Retrieving Microarray Experiments / 324
17.6 Similarity Metrics / 327
17.7 Evaluating Retrieval Performance / 329
17.8 Software Tools / 330
17.9 Conclusion and Future Directions / 331
Acknowledgment / 332
References / 332

18 EXTRACTION OF DIFFERENTIALLY EXPRESSED GENES IN MICROARRAY DATA 335
Tiratha Raj Singh, Brigitte Vannier, and Ahmed Moussa
18.1 Introduction / 335
18.2 From Microarray Image to Signal / 336
18.2.1 Signal from Oligo DNA Array Image / 336
18.2.2 Signal from Two–Color cDNA Array / 337
18.3 Microarray Signal Analysis / 337
18.3.1 Absolute Analysis and Replicates in Microarrays / 338
18.3.2 Microarray Normalization / 339
18.4 Algorithms for De Gene Selection / 339
18.4.1 Within Between DE Gene (WB–DEG) Selection Algorithm / 340
18.4.2 Comparison of the WB–DEGs with Two Classical DE Gene Selection Methods on Latin Square Data / 341
18.5 Gene Ontology Enrichment and Gene Set Enrichment Analysis / 343
18.6 Conclusion / 345
References / 345

19 CLUSTERING AND CLASSIFICATION TECHNIQUES FOR GENE EXPRESSION PROFILE PATTERN ANALYSIS 347
Emanuel Weitschek, Giulia Fiscon, Valentina Fustaino, Giovanni Felici, and Paola Bertolazzi
19.1 Introduction / 347
19.2 Transcriptome Analysis / 348
19.3 Microarrays / 349
19.3.1 Applications / 349
19.3.2 Microarray Technology / 350
19.3.3 Microarray Workflow / 350
19.4 RNA–Seq / 351
19.5 Benefits and Drawbacks of RNA–Seq and Microarray Technologies / 353
19.6 Gene Expression Profile Analysis / 356
19.6.1 Data Definition / 356
19.6.2 Data Analysis / 357
25.3.1 Implementation / 491
25.3.2 Synthetic Data / 491
25.3.3 Real Prokaryotic (Genomic) Data / 495
25.4 Results and Discussion / 501
25.4.1 Analysis of Synthetic Data / 501
25.4.2 Analysis of Prokaryotic Data / 502
25.5 Conclusion / 502
References / 503

VII PATTERN RECOGNITION IN BIOLOGICAL NETWORKS 505

26 COMPUTATIONAL METHODS FOR MODELING BIOLOGICAL INTERACTION NETWORKS 507
Christos Makris and Evangelos Theodoridis
26.1 Introduction / 507
26.2 Measures/Metrics / 508
26.3 Models of Biological Networks / 511
26.4 Reconstructing and Partitioning Biological Networks / 511
26.5 PPI Networks / 513
26.6 Mining PPI Networks Interaction Prediction / 517
26.7 Conclusions / 519
References / 519

27 BIOLOGICAL NETWORK INFERENCE AT MULTIPLE SCALES: FROM GENE REGULATION TO SPECIES INTERACTIONS 525
Andrej Aderhold, V Anne Smith, and Dirk Husmeier
27.1 Introduction / 525
27.2 Molecular Systems / 528
27.3 Ecological Systems / 528
27.4 Models and Evaluation / 529
27.4.1 Notations / 529
27.4.2 Sparse Regression and the LASSO / 530
27.4.3 Bayesian Regression / 530
27.4.4 Evaluation Metric / 531
27.5 Learning Gene Regulation Networks / 532
27.5.1 Nonhomogeneous Bayesian Regression / 533
27.5.2 Gradient Estimation / 534
27.5.3 Simulated Bio–PEPA Data / 534
27.5.4 Real mRNA Expression Profile Data / 535
27.5.5 Method Evaluation and Learned Networks / 536
27.6 Learning Species Interaction Networks / 540
27.6.1 Regression Model of Species interactions / 540
27.6.2 Multiple Global Change–Points / 541
27.6.3 Mondrian Process Change–Points / 542
27.6.4 Synthetic Data / 544
27.6.5 Simulated Population Dynamics / 544
27.6.6 Real World Plant Data / 546
27.6.7 Method Evaluation and Learned Networks / 546
27.7 Conclusion / 550
References / 550

28 DISCOVERING CAUSAL PATTERNS WITH STRUCTURAL EQUATION MODELING: APPLICATION TO TOLL–LIKE RECEPTOR SIGNALING PATHWAY IN CHRONIC LYMPHOCYTIC LEUKEMIA 555
Athina Tsanousa, Stavroula Ntoufa, Nikos Papakonstantinou, Kostas Stamatopoulos, and Lefteris Angelis

28.1 Introduction / 555
28.2 Toll–Like Receptors / 557
28.2.1 Basics / 557
28.2.2 Structure and Signaling of TLRs / 558
28.2.3 TLR Signaling in Chronic Lymphocytic Leukemia / 559
28.3 Structural Equation Modeling / 560
28.3.1 Methodology of SEM Modeling / 560
28.3.2 Assumptions / 561
28.3.3 Estimation Methods / 562
28.3.4 Missing Data / 562
28.3.5 Goodness–of–Fit Indices / 563
28.3.6 Other Indications of a Misspecified Model / 565
28.4 Application / 566
28.5 Conclusion / 580
References / 581

29 ANNOTATING PROTEINS WITH INCOMPLETE LABEL INFORMATION 585
Guoxian Yu, Huzefa Rangwala, and Carlotta Domeniconi
29.1 Introduction / 585
29.2 Related Work / 587
29.3 Problem Formulation / 589
29.3.1 The Algorithm / 591
29.4 Experimental Setup / 592
29.4.1 Data sets / 592
29.4.2 Comparative Methods / 593
29.4.3 Experimental Protocol / 594
29.4.4 Evaluation Criteria / 594
29.5 Experimental Analysis / 596
29.5.1 Replenishing Missing Functions / 596
29.5.2 Predicting Unlabeled Proteins / 600
29.5.3 Component Analysis / 604
29.5.4 Run Time Analysis / 604
29.6 Conclusions / 605
Acknowledgments / 606
References / 606
INDEX 609
Fax Order Form
To place an order via fax simply print this form, fill in the information below and fax the completed form to 646-607-1907 (from USA) or +353-1-481-1716 (from Rest of World). If you have any questions please visit http://www.researchandmarkets.com/contact/

Order Information
Please verify that the product information is correct.

<table>
<thead>
<tr>
<th>Product Name:</th>
<th>Pattern Recognition in Computational Molecular Biology. Techniques and Approaches. Wiley Series in Bioinformatics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web Address:</td>
<td><a href="http://www.researchandmarkets.com/reports/3335887/">http://www.researchandmarkets.com/reports/3335887/</a></td>
</tr>
<tr>
<td>Office Code:</td>
<td>SCDKLGJ5</td>
</tr>
</tbody>
</table>

Product Format
Please select the product format and quantity you require:

<table>
<thead>
<tr>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard Copy (Hard Back):</td>
</tr>
</tbody>
</table>

* Shipping/Handling is only charged once per order.

Contact Information
Please enter all the information below in BLOCK CAPITALS

<table>
<thead>
<tr>
<th>Title:</th>
<th>Mr</th>
<th>Mrs</th>
<th>Dr</th>
<th>Miss</th>
<th>Ms</th>
<th>Prof</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Name:</td>
<td>__________________________</td>
<td>Last Name:</td>
<td>__________________________</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email Address:</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job Title:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organisation:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Address:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>City:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postal / Zip Code:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phone Number:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fax Number:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Please refrain from using free email accounts when ordering (e.g. Yahoo, Hotmail, AOL)
Payment Information

Please indicate the payment method you would like to use by selecting the appropriate box.

☐ Pay by credit card: You will receive an email with a link to a secure webpage to enter your credit card details.

☐ Pay by check: Please post the check, accompanied by this form, to:
Research and Markets,
Guinness Center,
Taylors Lane,
Dublin 8,
Ireland.

☐ Pay by wire transfer: Please transfer funds to:
Account number 833 130 83
Sort code 98-53-30
Swift code ULSBIE2D
IBAN number IE78ULSB98533083313083
Bank Address Ulster Bank,
27-35 Main Street,
Blackrock,
Co. Dublin,
Ireland.

If you have a Marketing Code please enter it below:

Marketing Code: __________________________

Please note that by ordering from Research and Markets you are agreeing to our Terms and Conditions at http://www.researchandmarkets.com/info/terms.asp

Please fax this form to:
(646) 607-1907 or (646) 964-6609 - From USA
+353-1-481-1716 or +353-1-653-1571 - From Rest of World