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 researches in Computer Science, Life Science, and Mathematics. This book also serves as a supplementary reading for graduate students and young researches 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 Haplotype Problems 3
   Giuseppe Lancia
      1.1 Introduction / 3

5

1.2 Single Individual Haplotyping / 5
9 GENOME–WIDE SEARCH FOR PSEUDOKNOTTED NONCODING RNAs: A COMPARATIVE STUDY 155
Meghana Vasavada, Kevin Byron, Yang Song, and Jason T.L. Wang

9.1 Introduction / 155
9.2 Background / 156
9.2.1 Noncoding RNAs and Their Secondary Structures / 156
9.2.2 Pseudoknotted ncRNA Search Tools / 157
9.3 Methodology / 157
9.4 Results and Interpretation / 161
9.5 Conclusion / 162

References / 163

III PATTERN RECOGNITION IN TERTIARY STRUCTURES 165

10 MOTIF DISCOVERY IN PROTEIN 3D–STRUCTURES USING GRAPH MINING TECHNIQUES 167
Wajdi Dhifli and Engelbert Mephu Nguifo

10.1 Introduction / 167
10.2 From Protein 3D–Structures to Protein Graphs / 169
10.2.1 Parsing Protein 3D–Structures into Graphs / 169
10.3 Graph Mining / 172
10.4 Subgraph Mining / 173
10.5 Frequent Subgraph Discovery / 173
10.5.1 Problem Definition / 174
10.5.2 Candidates Generation / 176
10.5.3 Frequent Subgraph Discovery Approaches / 177
10.5.4 Variants of Frequent Subgraph Mining: Closed and Maximal Subgraphs / 178
10.6 Feature Selection / 179
10.6.1 Relevance of a Feature / 179
10.7 Feature Selection for Subgraphs / 180
10.7.1 Problem Statement / 180
10.7.2 Mining Top–k Subgraphs / 180
10.7.3 Clustering–Based Subgraph Selection / 181
10.7.4 Sampling–Based Approaches / 181
14.1 Introduction / 251
14.2 Protein Structure Prediction / 255
14.2.1 Secondary Structure Prediction / 255
14.2.2 Modeling of Tertiary Structure / 256
14.3 Template–Based Predictions / 257
14.3.1 Homology Modeling / 257
14.3.2 Threading Methods / 257
14.3.3 Ab initio Modeling / 257
14.4 Critical Assessment of Protein Structure Prediction / 258
14.5 Quaternary Structure Prediction / 258
14.6 Conclusion / 261
Acknowledgments / 261
References / 261

15 COMPARISON OF PROTEIN QUATERNARY STRUCTURES BY GRAPH APPROACHES 266
Sheng–Lung Peng and Yu–Wei Tsay
15.1 Introduction / 266
15.2 Similarity in the Graph Model / 268
15.2.1 Graph Model for Proteins / 270
15.3 Measuring Structural Similarity VIA MCES / 272
15.3.1 Problem Formulation / 273
15.3.2 Constructing P–Graphs / 274
15.3.3 Constructing Line Graphs / 276
15.3.4 Constructing Modular Graphs / 276
15.3.5 Maximum Clique Detection / 277
15.3.6 Experimental Results / 277
15.4 Protein Comparison VIA Graph Spectra / 279
15.4.1 Graph Spectra / 279
15.4.2 Matrix Selection / 281
15.4.3 Graph Cospectrality and Similarity / 283
15.4.4 Cospectral Comparison / 283
15.4.5 Experimental Results / 284
15.5 Conclusion / 287
16 STRUCTURAL DOMAINS IN PREDICTION OF BIOLOGICAL PROTEIN PROTEIN INTERACTIONS

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

17 CONTENT-BASED RETRIEVAL OF MICROARRAY EXPERIMENTS

Hasan Oğul

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
24 AUTOMATED PLAUSIBILITY ANALYSIS OF LARGE PHYLOGENIES 457
David Dao, Tomáš Flouri, and Alexandros Stamatakis

24.1 Introduction / 457
24.2 Preliminaries / 459
24.3 A Naïve Approach / 462
24.4 Toward a Faster Method / 463
24.5 Improved Algorithm / 467
24.5.1 Preprocessing / 467
24.5.2 Computing Lowest Common Ancestors / 468
24.5.3 Constructing the Induced Tree / 468
24.5.4 Final Remarks / 471
24.6 Implementation / 473
24.6.1 Preprocessing / 473
24.6.2 Reconstruction / 473
24.6.3 Extracting Bipartitions / 474
24.7 Evaluation / 474
24.7.1 Test Data Sets / 474
24.7.2 Experimental Results / 475
24.8 Conclusion / 479

Acknowledgment / 481
References / 481

25 A NEW FAST METHOD FOR DETECTING AND VALIDATING HORIZONTAL GENE TRANSFER EVENTS USING PHYLOGENETIC TREES AND AGGREGATION FUNCTIONS 483
Dunarel Badescu, Nadia Tahiri, and Vladimir Makarenkov

25.1 Introduction / 483
25.2 Methods / 485
25.2.1 Clustering Using Variability Functions / 485
25.2.2 Other Variants of Clustering Functions Implemented in the Algorithm / 487
25.2.3 Description of the New Algorithm / 488
25.2.4 Time Complexity / 491
25.3 Experimental Study / 491
### 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
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.

Product Name: Pattern Recognition in Computational Molecular Biology. Techniques and Approaches. Wiley Series in Bioinformatics
Web Address: http://www.researchandmarkets.com/reports/3335887/
Office Code: SCBRK8QZ

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

| Quantity          | Hard Copy (Hard Back): | USD 127 + USD 29 Shipping/Handling |

* Shipping/Handling is only charged once per order.

Contact Information
Please enter all the information below in BLOCK CAPITALS

Title: Mr ☐ Mrs ☐ Dr ☐ Miss ☐ Ms ☐ Prof ☐
First Name: __________________________ Last Name: __________________________
Email Address: * __________________________
Job Title: __________________________
Organisation: __________________________
Address: __________________________
City: __________________________
Postal / Zip Code: __________________________
Country: __________________________
Phone Number: __________________________
Fax Number: __________________________

* 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 IE78ULSB9853308313083
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