
DATA MINING IN BIOMEDICINE

Edited by

PANOS M. PARDALOS

University of Florida, Gainesville, FL

VLADIMIR L. BOGINSKI

Florida State University, Tallahassee, FL

ALKIS VAZACOPOULOS

Dash Optimization, Englewood Cliffs, NJ

 Springer

Library of Congress Control Number: 2006938889

ISBN-10: 0-387-69318-1

e-ISBN-10: 0-387-69319-X

ISBN-13: 978-0-387-69318-7

e-ISBN-13: 978-0-387-69319-4

Printed on acid-free paper.

AMS Subject Classifications: 65K05, 90C90, 62-07

© 2007 Springer Science+Business Media, LLC

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

springer.com

Contents

Preface	ix
List of Contributors	xi

Part I Recent Methodological Developments for Data Mining Problems in Biomedicine

Pattern-Based Discriminants in the Logical Analysis of Data <i>Sorin Alexe, Peter L. Hammer</i>	3
Exploring Microarray Data with Correspondence Analysis <i>Stanislav Busygin, Panos M. Pardalos</i>	25
An Ensemble Method of Discovering Sample Classes Using Gene Expression Profiling <i>Dechang Chen, Zhe Zhang, Zhenqiu Liu, Xiuzhen Cheng</i>	39
CpG Island Identification with Higher Order and Variable Order Markov Models <i>Zhenqiu Liu, Dechang Chen, Xue-wen Chen</i>	47
Data Mining Algorithms for Virtual Screening of Bioactive Compounds <i>Mukund Deshpande, Michihiro Kuramochi, George Karypis</i>	59
Sparse Component Analysis: a New Tool for Data Mining <i>Pando Georgiev, Fabian Theis, Andrzej Cichocki, Hovagim Bakardjian</i> ..	91
Data Mining Via Entropy and Graph Clustering <i>Anthony Okafor, Panos Pardalos, Michelle Ragle</i>	117

Molecular Biology and Pooling Design <i>Weili Wu, Yingshu Li, Chih-hao Huang, Ding-Zhu Du</i>	133
An Optimization Approach to Identify the Relationship between Features and Output of a Multi-label Classifier <i>Musa Mammadov, Alex Rubinov, John Yearwood</i>	141
Classifying Noisy and Incomplete Medical Data by a Differential Latent Semantic Indexing Approach <i>Liang Chen, Jia Zeng, Jian Pei</i>	169
Ontology Search and Text Mining of MEDLINE Database <i>Hyunki Kim, Su-Shing Chen</i>	177
<hr/>	
Part II Data Mining Techniques in Disease Diagnosis	
<hr/>	
Logical Analysis of Computed Tomography Data to Differentiate Entities of Idiopathic Interstitial Pneumonias <i>M.W. Brauner, N. Brauner, P.L. Hammer, I. Lozina, D. Valeyre</i>	193
Diagnosis of Alport Syndrome by Pattern Recognition Techniques <i>Giacomo Patrizi, Gabriella Addonisio, Costas Giannakakis, Andrea Onetti Muda, Gregorio Patrizi, Tullio Faraggiana</i>	209
Clinical Analysis of the Diagnostic Classification of Geriatric Disorders <i>Giacomo Patrizi, Gregorio Patrizi, Luigi Di Cioccio, Claudia Bauco</i>	231
<hr/>	
Part III Data Mining Studies in Genomics and Proteomics	
<hr/>	
A Hybrid Knowledge Based-Clustering Multi-Class SVM Approach for Genes Expression Analysis <i>Budi Santosa, Tyrrell Conway, Theodore Trafalis</i>	261
Mathematical Programming Formulations for Problems in Genomics and Proteomics <i>Cláudio N. Meneses, Carlos A.S. Oliveira, Panos M. Pardalos</i>	275
Inferring the Origin of the Genetic Code <i>Maria Luisa Chiusano, Luigi Frusciante, Gerardo Toraldo</i>	291
Deciphering the Structures of Genomic DNA Sequences Using Recurrence Time Statistics <i>Jian-Bo Gao, Yinhe Cao, Wen-wen Tung</i>	321

Clustering Proteomics Data Using Bayesian Principal Component Analysis
Halima Bensmail, O. John Semmes, Abdelali Haoudi 339

Bioinformatics for Traumatic Brain Injury: Proteomic Data Mining
Su-Shing Chen, William E. Haskins, Andrew K. Ottens, Ronald L. Hayes, Nancy Denslow, Kevin K.W. Wang 363

Part IV Characterization and Prediction of Protein Structure

Computational Methods for Protein Fold Prediction: an Ab-initio Topological Approach
G. Ceci, A. Mucherino, M. D'Apuzzo, D. Di Serafino, S. Costantini, A. Facchiano, G. Colonna 391

A Topological Characterization of Protein Structure
Bala Krishnamoorthy, Scott Provan, Alexander Tropsha 431

Part V Applications of Data Mining Techniques to Brain Dynamics Studies

Data Mining in EEG: Application to Epileptic Brain Disorders
W. Chaovalitwongse, P.M. Pardalos, L.D. Iasemidis, W. Suharitdamrong, D.-S. Shiau, L.K. Dance, O.A. Prokopyev, V.L. Boginski, P.R. Carney, J.C. Sackellares 459

Information Flow in Coupled Nonlinear Systems: Application to the Epileptic Human Brain
S. Sabesan, K. Narayanan, A. Prasad, L. D. Iasemidis, A. Spanias, K. Tsakalis 483

Reconstruction of Epileptic Brain Dynamics Using Data Mining Techniques
Panos M. Pardalos, Vitaliy A. Yatsenko 505

Automated Seizure Prediction Algorithm and its Statistical Assessment: A Report from Ten Patients
D.-S. Shiau, L.D. Iasemidis, M.C.K. Yang, P.M. Pardalos, P.R. Carney, L.K. Dance, W. Chaovalitwongse, J.C. Sackellares 517

Seizure Predictability in an Experimental Model of Epilepsy
S.P. Nair, D.-S. Shiau, L.D. Iasemidis, W.M. Norman, P.M. Pardalos, J.C. Sackellares, P.R. Carney 535

**Network-Based Techniques in EEG Data Analysis and
Epileptic Brain Modeling**

*Oleg A. Prokopyev, Vladimir L. Boginski, Wanpracha Chaovaitwongse,
Panos M. Pardalos, J. Chris Sackellares, Paul R. Carney* 559

Index 575