

CONFERENCE PROGRAM FOR FEBRUARY 2, 2005 Wednesday (282 Reitz Union)			
8:00	Registration		
8:45 - 9:00	Welcome Session Panos M. Pardalos		
9:00 - 10:30	Session W1 Chair: Panos Pardalos		
	Theodore Trafalis	University of Oklahoma	Maximum margin classifiers with noisy data: A robust optimization approach
	Gary Kochenberger Fred Glover	University of Colorado	Improved Clustering of Microarray Data for Genomics
	Carlos Oliveira	Oklahoma State University	Algorithmic Results for the Maximum Likelihood Problem on Evolutionary Trees
10:30 - 10:45	Coffee Break		
10:45 - 12:15	Session W2 Chair: Stanislav Busygin		
	David Halitsky	Cumulative Inquiry, Inc.	A conserved sparse dicodon framework which correlates sequence and structure: implications for gene finding
	Michael Baudis	University of Florida	Large-scale cytogenetic data parsing for oncogenomic gene and pathway identification
	Wanpracha Chaovalitwongse	Rutgers University	A Robust Combinatorial Approach for Sibling Relationship Reconstruction
12:15 - 1:30	Lunch		
1:30 - 3:00	Session W3 Chair: Maria Luisa Chiusano		
	Maria Luisa Chiusano	University of Naples "Federico II", Italy	Mining integrated data to investigate relationships between nucleic acids and proteins
	Giacomo Patrizi	University "La Sapienza", Rome Italy	Solving large Protein Folding Problems by a linear complementarity algorithm with $\{0,1\}$ variables
	Antonio Mucherino	Second University of Naples, Italy	Analysing protein structural datasets to get information on protein shapes
3:00 - 3:30	Coffee Break		
3:30 - 5:00	Session W4 Chair: Paul Carney		
	J. Chris Sackellares	University of Florida	The chaotic brain: nonlinear dynamics and epilepsy
	Paul Carney	University of Florida	Dynamical Characteristics in the Rat Chronic Limbic Epilepsy Model
	Andrew Ottens	University of Florida	Data mining and bioinformatic analysis in neuroproteomics of brain injury
6:30	Dinner (Hilton Hotel)		

CONFERENCE PROGRAM FOR FEBRUARY 3, 2005 Thursday (282 Reitz Union)

8:30 - 10:30	Session R1 Chair: Panos Pardalos		
	Stanislav Busygin	University of Florida	Optimization framework for feature selection, classification and outlier detection
	Mingzhou Ding	University of Florida	Analyzing Multichannel Neural Data: A Comprehensive Framework
	Musa Mammadov	University of Ballarat, Australia	The study of drug-reaction association using global optimization techniques
	Pando Georgiev	University of Cincinnati	A Sparse Representation Algorithm Based on Global Optimization Technique
10:30 - 10:45	Coffee Break		
10:45 - 12:15	Session R2 Chair: Vladimir Boginski		
	Sergiy Butenko	Texas A&M University	Analyzing the Structure of Biological Networks
	Nikos Tsoukias	Florida International University	Multiscale models of NO and O ₂ biotransport in the microcirculation
	Rongling Wu	University of Florida	An emerging conceptual framework of cancer gene identification
12:15 - 1:30	Lunch		
1:30 - 3:00	Session R3 Chair: TBA		
	Deng-Shan Shiau	University of Florida	Pattern-match regularity statistics - A measure quantifying the characteristics of epileptic seizures
	Zhiqun Zhang	University of Florida	Systematic approach to decipher neuritogenesis: ROCK pathways in mediating neurite outgrowth in PC-12 cells
	Sung-Phil Kim	University of Florida	Real-time selection of subset of neurons in BMIs
3:00 - 3:30	Coffee Break		
3:30 - 5:00	Session R4 Chair: TBA		
	Xing Liu	University of Maryland, Baltimore County	Aligning families of 2D-Gels by a combined multiresolution forward-inverse transformation approach
	Jianbo Gao	University of Florida	Novel unsupervised methods for identifying protein coding sequence from genomic DNA sequences
	Ding-Zhu Du	National Science Foundation	Pooling Designs in Molecular Biology

CONFERENCE PROGRAM FOR FEBRUARY 4, 2005 Friday (282 Reitz Union)

9:00 - 10:30	Session F1		
	Chair: TBA		
	Nasir Hussain	Univ. of Delaware	Statistical analysis of baroreceptors in rabbits
	Arunava Banerjee	University of Florida	The Spike Activity of Neocortical Columns: A Dynamical Systems Analysis
	Sergei Pilyugin	University of Florida	Rescaling method for quantifying the turnover rates of lymphocytes using the CFSE assay
10:30 - 10:45	Coffee Break		
10:45 - 12:15	Session F2		
	Chair: TBA		
	H. Edwin Romeijn	University of Florida	A column generation approach to radiation therapy treatment planning using aperture modulation
	M. Davidson	University of Florida	Directions in Analysis of Large Data Sets from Synchrotron X-ray Studies of Tissue Samples
	Dong Hua	George Washington University	Single digit region of interest (SDROI) extraction by reducing the tissue influence
12:15 - 1:30	Lunch		
1:30 - 3:00	Session F3		
	Chair: TBA		
	Bernard A. Mair	University of Florida	Simultaneous Estimation of Myocardial Intensity and Motion
	Cristian Cardenas-Lailhacar	University of Florida	Capacitance, hardness and the potential energy function: A model for chemical reactivity prediction
	Claudio Meneses	University of Florida	Modeling and Solving String Selection Problems
3:00 - 3:30	Coffee Break		
3:30 - 5:00	Session F4		
	Chair: TBA		
	Vitaliy Yatsenko	Institute of Space Research, Ukraine	Control of Lyapunov Exponents and Biomedical Applications
	Cheong-Hee Park	University of Florida	Quality assessment in feature selection
	P.I. Stetsyuk	V.M. Glushkov Institute of Cybernetics, Ukraine	Quadratic-type models and upper bounds for the problems of finding the maximum weighted independent set in graphs

Maximum Margin Classifiers with Noisy Data: A Robust Optimization Approach

Theodore B. Trafalis

Laboratory of Optimization and Intelligent Systems
School of Industrial Engineering
University of Oklahoma 202 West Boyd, Room 124, Norman, OK 73019

ttrafalis@ou.edu

Abstract

A robust convex optimization approach is proposed for maximum margin classifiers with noisy input data. This approach deals with training data that belong to convex uncertainty sets; what is known about the uncertain data vector v , referring to a perturbation of an input data point x , is that it belongs to a given uncertainty set U defined as the sphere of an L_p norm. We show that in the case that the set U is an L_2 sphere the resulting optimization problem is equivalent to a Second Order Cone Programming (SOCP) Problem. Formulations based on L_1 and L_8 norms are also examined. Applications to biomedical data are discussed.

Keywords: Optimization, Robustness, Parameter estimation, Kernel Functions, Support Vector Machines, Second order cone programming, Maximum Margin Classifiers

Improved Clustering of Microarray Data for Genomics

Gary Kochenberger¹, Fred Glover², Bahram Alidaee³, Haibo Wang³

¹University of Colorado at Denver

²University of Colorado at Boulder

³University of Mississippi

Gary.Kochenberger@Cudenver.edu

Abstract

An enhanced understanding of functional genomics results from data mining aimed at discovering hidden patterns. The availability of microarray data creates new opportunities for this. An important early step in such analysis is to apply clustering techniques to reveal structures and patterns not otherwise known or confirmed in the underlying data. We present a new model for cluster analysis based on clique partitioning that holds great promise for analyzing high dimensional gene expression data. Preliminary computational experience with several data sets shows the efficacy of the model.

Algorithmic Results for the Maximum Likelihood Problem on Evolutionary Trees

Carlos A.S. Oliveira¹, Panos M. Pardalos²

¹School of Industrial Engineering and Management, Oklahoma State University,
322 Engineering North, Stillwater, OK 74078.

²Center for Applied Optimization, Department of Industrial and Systems Engineering,
University of Florida. Address: 303 Weil Hall, Gainesville, FL 32611.

coliv@okstate.edu

Abstract

Maximum likelihood (ML) is a problem arising in the reconstruction of evolutionary trees, derived from the analysis of genomic data. ML has been used by several researchers due to the good results provided by the corresponding selection of evolutionary trees. However, the general problem arising from ML is computationally difficult to solve, especially when compared to competing techniques such as maximum parsimony (MP). In the ML problem we are given a set of m sequences of characters, each with the same length n . The objective of the problem is to find a tree T with m leaves, an assignment of probabilities $p : E(T) \rightarrow [0, 1]$, and a labeling $\lambda : V(T) \rightarrow 0, 1^n$ such that the resulting probability of occurrence of tree T is maximized. Some restricted versions of the ML are solvable in polynomial time, however it is believed that the ML, as occurs with the MP, is NP-hard in general. In this talk, we present results concerning the complexity of the ML, and propose some algorithmic strategies for solving this problem.

A Conserved Sparse Dicodon Framework Which Correlates Sequence and Structure: Implications for Gene Finding

David J. Halitsky¹, Arthur M. Lesk², Jacques R. Fresco³

¹Cumulative Inquiry, Com., Thompsons Station, TN

²Department of Haematology, CIMR, Cambridge, UNITED KINGDOM

³Department of Molecular Biology, Princeton University, Princeton, NJ

Abstract

Analysis of di-codon pairs in mRNA sequences can identify regions of structural similarity in the encoded proteins. The signal is only sparsely distributed in mRNAs, but we have written algorithms that distinguish signal reliably from noise. The patterns appear to be sensitive directly to sequence-structure correlations rather than evolutionary relationship, because they link regions in non-homologous proteins as well as detect related proteins. In this respect, the signal patterns are quite different from, and complementary to, other methods of searching for sequence-structure correlations. For example, the signal is strong enough that two manifestations of the same signal pattern P in the coding sequences for two entirely unrelated proteins A and B (not necessarily in the PDB) can be used conjointly to search successfully for cases in which the pattern P identifies coding subsequences underlying structurally related regions of two other proteins C and D in the PDB. The growing set of illustrative examples at the StrucClues website <http://www.CumulativeInquiry.com/StrucClues> shows that such structurally related regions may manifest tertiary and quaternary, as well as supersecondary, similarities. Further, this set of examples indicates that the proposed signal can be used for gene finding without recourse to any degree of sequence similarity.

Large-scale cytogenetic data parsing for oncogenomic gene and pathway identification

Michael Baudis

Div. of Pediatric Hematology/Oncology and Shands Cancer Center
University of Florida, Gainesville

Abstract

Malignant neoplasias rely on the abnormal expression of genes involved in organized cell death, growth and differentiation. Aiming at the identification of cancer related genes, the genomic composition of tens of thousands malignant tumors has been analyzed. Until recently, the vast majority of oncogenomic screening experiments was performed using chromosomal analysis techniques or molecular-cytogenetic techniques in a chromosomal reference space (e.g. Comparative Genomic Hybridization, CGH). Though these studies have led to important insights into the genetic mechanisms of tumor development, the idiosyncratic, chromosome based annotation format in cytogenetic reports prohibits automatic large-scale data mining efforts.

The Progenetix.net (molecular-) cytogenetic workbench is being developed as central online resource for collection and integration of oncogenomic data. Unique, dedicated parsing algorithms perform the transformation of ISCN related karyotype annotations into an interval specific status matrix, suitable for matching to other genome mapped information (e.g. from expression array experiments). Additionally to the life online conversion of user supplied cytogenetic data, the Progenetix website provides with more than 10000 cases the largest collection of Comparative Genomic Hybridization (CGH) cases and a trial version of the pre-formatted cases from the Mitelman collection.

As part of the Progenetix project, tools for visualization and analysis of oncogenomic data are being explored. Currently, the website allows for band specific status searches, generation of overview ideograms, clustering of tumors according to their aberration patterns and direct search for tumor entities with the highest aberration score for any given band. For data exchange, a custom XML format has been developed.

Access to band interval matched status data from currently more than 55000 human cancer and leukemia cases should provide a solid foundation for a systematic analysis of aberration patterns for pathway description, as well as for the interpretation of complex results from expression array experiments. Currently, advanced statistical methods are being explored for the description of oncogenomic data structures.

A Robust Combinatorial Approach for Sibling Relationship Reconstruction

Wanpracha Art Chaovalitwongse

Department of Industrial and Systems Engineering, Rutgers University
DIMACS, Rutgers University

Abstract

We present a novel combinatorial approach for reconstructing sibling relationships in a single generation of individuals without parental information, using data from codominant DNA markers such as microsatellites. We use the simple genetic constraints on the full-sibling groups, imposed by the Mendelian inheritance rules, and combinatorial optimization techniques to extract a minimum number of consistent sibling groups. The results of a simulation study of a relaxed version of the algorithm show that our approach is reasonably accurate and the full version of the algorithm should be pursued. Our algorithm does not require any a priori knowledge about allele frequency, population size, mating system, or family size distributions.

Mining integrated data to investigate relationships between nucleic acids and proteins

Maria Luisa Chiusano¹, Luigi Frusciante², Gerardo Toraldo³

¹Department of Genetics, General and Molecular Biology,
University of Naples "Federico II", via Mezzocannone 8, 80134 Naples, Italy

²Department of Soil, Plant and Environmental Sciences (DISSPA)
University of Naples "Federico II", via Universit 100, 80055 Portici, Italy

³Department of Agricultural Engineering
University of Naples "Federico II", via Universit 100, 80055 Portici , Italy

"chiusano@unina.it"

Abstract

The wide production of data concerning structural and functional aspects of molecules of fundamental biological interest of the last 30 years, mainly due to the rapid evolving of biotechnologies as well as to the accomplishment of the Genome Projects, led to the necessity of appropriate computational approaches for data storage, manipulation and analyses, giving place to fast evolving areas of Biology: Computational Biology and Bioinformatics. Because of the strong support that the collected data can provide to the biological knowledge, the design of methodologies suitable to help in the extraction of information hidden in the data became one of the major challenges in this area of Biology, with applications of strong impact in biology, medicine and biotechnologies. We discuss here about an integrated methodology to analyze the relationships between nucleic acids and protein structures by a comprehensive approach. Data deriving from both sequence and higher order structure information of both types of molecules were considered to produce an integrated data warehouse for supporting data mining based both on graphical approaches and on suitable statistics to extract biologically relevant information. To remark the usefulness of such a methodology, we report on specific applications to the study of relationships between coding regions and the corresponding encoded protein structures that revealed novel information in biology in the light of understanding structural, functional and evolutionary relationships (Chiusano et al., 1999; Chiusano, 2000; Chiusano et al., 2001; Facchiano et al., 2001; Potenza et al., 2003).

*Solving large Protein Folding Problems by a linear complementarity algorithm
with $\{0,1\}$ variables*

Claudio Cifarelli¹, Giacomo Patrizi²

¹ Dept. of Industrial and Systems Engineering,
University of Florida, Gainesville Fl.

² Dipartimento di Statistica, Probabilità e Statistiche Applicate
Università "La Sapienza", Rome Italy

"patrizi@banach.sta.uniroma1.it"

Abstract

Classification algorithms may be formulated as 0,1 integer programming problems or solved by an iterative algorithm. T.R.A.C.E. (Total Recognition by Adaptive Classification Experiments) which has proved efficient. The iterative version of the algorithm, can be adapted easily to an optimization version, formulated as a linear complementarity problem. The aim of this paper is to present a linear complementary problem algorithm, limited to 0,1 variables, determine its properties and give extensive application results by solving classification problems regarding protein folds and compare the results to those of the iterative algorithm version and other implementations. The increase in the recognition accuracy that will be found is due to the algorithm used, which allows the inclusion of constraints to guide the recognition and to the strict statistical methodology that is followed.

Analysing protein structural datasets to get information on protein shapes

A. Mucherino^{1,5}, M. D'Apuzzo^{1,4,5}, D. di Serafino^{1,4,5}, S. Costantini^{2,3,5}, A. Facchiano^{3,4,5} and G. Colonna^{2,4,5}

¹ Department of Mathematics, Second University of Naples, via Vivaldi 43, I-81100 Caserta, Italy.

² Department of Biochemistry and Biophysics, Second University of Naples, via Costantinopoli 16, I-80138 Naples, Italy.

³ Institute of Food Science, CNR, via Roma 52 A/C, I-83100 Avellino, Italy.

⁴ Research Center of Computational and Biotechnological Sciences (CRISCEB), Second University of Naples, via Costantinopoli 16, I-80138 Naples, Italy.

⁵ Computational Biology Doctorate, Second University of Naples, Italy.

Abstract

The prediction of protein folds can benefit from information on "geometrical" features of the proteins, such as shape and surface. The proteins may assume different shapes, i.e. globular, fibrous rod and others for not soluble proteins. Globular proteins are widely studied and many globular conformations have been experimentally solved; however, the laws that determine their shape and fold are still unknown. It is widely believed that the most important forces involved in the folding process are the hydrophobic and Van der Waals ones. In some studies, the hydrophobic forces have been related to the protein shape and to the protein surface, in order to find some rules that the protein three-dimensional conformations must obey. We present a method for predicting the shape of globular proteins. The shape of the target protein is modelled as an ellipsoid and the semiaxis lengths of the ellipsoid that better approximate this shape are searched. By analysing a set of proteins with known structure, some information is extracted that allows to correlate the hydrophobicity of the protein amino acids with the volume of the protein hydrophobic core, the thickness of the hydrophilic protein shell and the protein surface. This information is used to select a small number of semiaxis triplets, that are good candidates to determine the protein shape. A software tool has been developed to perform this analysis and has been applied to a protein set available from PDBselect. The obtained results show that, given the sequence of amino acid residues of a protein, the method is able to choose few triplets of semiaxis lengths, among which there is one corresponding to a good approximation of the protein shape. This procedure can be used in an ab-initio topological approach for protein fold prediction that requires information on the protein shape.

The Chaotic Brain: Nonlinear Dynamics and Epilepsy

J. Chris Sackellares

Department of Biomedical Engineering, University of Florida

Abstract

Although electroencephalographic (EEG) signal appears to offer only a one dimensional view of brain electrical activity, such a time series can provide information about a large number of pertinent variables. These variables and their evolution in time produce the phase portrait of the system. Such information may be used to explore and characterize the dynamics of the underlying system. EEG recordings from chronically implanted electrodes were analyzed. The existence of an attractor (convergence of the phase trajectories towards a subset of the phase space), its dimension, as well as proof of its chaotic nature (at least one positive Lyapunov exponent) in epileptic EEG signals will be demonstrated. Data from the preictal and ictal phases of epileptic seizures provide a unique opportunity to follow the variation in time of dimensionality and chaoticity of the attractors. These results can be used in the detection and prediction of the onset of a seizure and in the parametric modeling of the phenomenon itself, which constitute active areas of research today.

Dynamical Characteristics in the Rat Chronic Limbic Epilepsy Model

Paul R. Carney

Department of Pediatrics, University of Florida

Abstract

We have previously reported preictal transitions, detectable in the spatiotemporal characteristics of the EEG signal in human mesial temporal lobe epilepsy (MTLE) using short term largest Lyapunov exponent (STLmax) and average angular frequency ($\bar{\Omega}$). These results have prompted us to apply the quantitative nonlinear methods to a chronic limbic epilepsy rat (CLE) model, as this model has several important features of human MTLE. The present study tests the hypothesis that preictal dynamical changes, similar to those observed in humans exist in the CLE model. Twenty-eight, 2-hr epoch data sets from 4 CLE rats (mean seizure duration 7821 sec) are analyzed, each containing a focal onset seizure and intracranial data beginning 1 hr before the seizure onset. The signal was tested for presence of nonlinearities using the correlation integral measure and comparing with surrogates datasets. Analysis of STLmax shows multiple transient drops in STLmax values during the preictal period followed by a significant drop during the ictal period. Average angular frequency values demonstrate transient peaks during the preictal period followed by a significant peak during the ictal period. Convergence among electrode sites is also observed in both STLmax and $\bar{\Omega}$ values before seizure onset.

Results indicate that there are characteristic spatiotemporal changes in the EEG signal that precede and accompany seizures in rat CLE. Thus, it may be possible to use the rat CLE model to investigate dynamical mechanisms underlying seizure onset and cessation, and as a tool to refine and test real-time seizure detection prediction, and closed-loop intervention techniques.

Data Mining and Bioinformatic Analysis in Neuroproteomics of Brain Injury

**Andrew K. Ottens^{1,2,3}, Firas H. Kobeissy^{1,2,4}, Zhiqun Zhang^{1,2,3}, Erin C. Golden^{1,3},
Regina A. Wolper^{1,3}, Ming Chen Liu^{1,2,3}, Stanislav Busygin⁵, Alla Revenko⁵, Panos M.
Pardalos⁵, Su-Shing Chen⁶, Ronald L. Hayes^{2,3,4}, Kevin K. W. Wang^{1,2,3,4}**

Centers for Neuroproteomics and Biomarkers Research¹ and Traumatic Brain Injury Studies²;
Departments of Neuroscience³ and Psychiatry⁴ of the McKnight Brain Institute; Department of
Industrial and Systems Engineering⁵; Department of Computer and Information Science and
Engineering⁶; University of Florida, Gainesville, FL 32611.

Abstract

Bioanalytical analysis verges on complexity only rivaled by its target biological system. Flurries of techniques prod away at hidden intricacies underlying the proteome, an assemblage of proteins with their modifications and interactions that govern our thoughts, feelings, and actions at the molecular level. In the brain, the neuroproteome reveals how cells respond to injury, what damage is made, and ultimately how we recover lost function. Tools at hand are beginning to tap into this unseen world and are providing promising diagnostic and putatively therapeutic avenues. Yet as exploration is ramped up, the research laboratory is overwhelmed with immense amounts of seemingly unconnected data. Here we present the challenge applied to brain injury neuroproteomics. We explore the biological relationship between data, and outline the informatic platform from a scientist perspective needed to unfold key discoveries in neuroproteomics.

Optimization framework for feature selection, classification and outlier detection

Stanislav Busygin, Oleg Prokopyev, Panos M. Pardalos

Industrial and Systems Engineering
University of Florida
303 Weil Hall, Gainesville, FL 32611

busygin@ufl.edu

Abstract

Let a data set of n samples and m features be given as a rectangular matrix $A = (a_{ij})_{m \times n}$, where the value a_{ij} is the expression of i -th feature in j -th sample. The set of samples is divided into the *training* and *test* sets. The samples from the training set are partitioned into r classes C_1, C_2, \dots, C_r , while classification of the test samples is unknown. The problem is to classify the test samples into the given r classes provided the known classification of the training samples. This is one of the central problems of data mining theory and applications, and in practice it is frequently complicated by the presence of *outliers* (i.e., samples which do not possess characteristics of the majority of samples from their class) in the training set. Furthermore, usually not all features of the data are informative for discovering the classification, and a subset of features determining it should be found. This task is called the *feature selection*.

We consider several unsupervised clustering principles and use them as constraints for the feature selection. Introducing a vector of variables $x = (x_i)_{i=1..m}$ for feature weights bounded between 0 and 1, we can formulate the constraints

$$\sum_{i=1}^m \left(a_{ij} - c_i^{(\hat{k})} \right)^2 x_i \leq \sum_{i=1}^m \left(a_{ij} - c_i^{(k)} \right)^2 x_i$$

to impose (local) optimality of the given classification for k -means clustering (where $c^{(\hat{k})}$ is the centroid of the class to which the j -th sample belongs, and $c^{(k)}$ is the centroid of any other class); the constraints

$$\sum_{i=1}^m (a_{ij_1} - a_{ij_2})^2 x_i \leq \sum_{i=1}^m (a_{ij_3} - a_{ij_4})^2 x_i,$$

whenever the samples j_1 and j_2 are from the same class, and samples j_3 and j_4 are from different classes, to impose global optimality of the given classification.

Next, we may construct classification of features corresponding to the given classification of samples assigning the i -th feature to class $C_{\hat{k}}$ if $c_i^{(\hat{k})} = \max_k c_i^{(k)}$. We consider the *biclustering consistency criterion* asserting that “backward” classification of samples based on the classification of features must be the original given classification: if the j -th sample belongs to class $C_{\hat{k}}$, then

$$\frac{\sum_{i=1}^m a_{ij} f_{i\hat{k}} x_i}{\sum_{i=1}^m f_{i\hat{k}} x_i} \geq \frac{\sum_{i=1}^m a_{ij} f_{ik} x_i}{\sum_{i=1}^m f_{ik} x_i},$$

where $f_{ik} = 1$ if the i -th feature is classified to class C_k , and $f_{ik} = 0$ otherwise.

Having such constraints originating from unsupervised clustering criteria, we may optimize a certain objective function either minimizing information loss incurred due to the feature space reduction or maximizing the separation between classes. Then, the classification for the test samples is performed with respect to the same clustering criterion that was used for the feature selection.

When the training set contains outliers, the quality of the obtained solution becomes not satisfactory, but analyzing which constraints give Lagrangian multipliers (dual variables) of highest magnitude, we are able to detect what samples are outliers in each of the classes. Removing some or all constraints involving them, we will improve the quality of the solution.

The developed optimization-based approach has shown good performance on well-known DNA microarray data sets.

Analyzing Multichannel Neural Data: A Comprehensive Framework

Mingzhou Ding

Department of Biomedical Engineering
149 BME Building/PO Box 116131
University of Florida, Gainesville, FL 32611

`mding@bme.ufl.edu`

Abstract

The advance of experimental technology is making simultaneous multichannel neural recordings a daily reality in labs around the world. How to process the massive quantities of data to gain useful insights into the workings of neurobiological systems is a key issue facing neuroscientists today. In this talk I will discuss recent progress in this area in the context of monkeys performing cognitive tasks.

The study of Drug-reaction Associations Using Global Optimization Techniques

Mammadov M.A., Rubinov A.M. and J.Yearwood

School of Information Technology and Mathematical Sciences
University of Ballarat Victoria, 3353, Australia

Abstract

The accurate identification of drugs that are responsible for reactions that have occurred is one of the important problems of adverse drug reactions (ADR). In this paper we formulate this problem as a multi-label classification problem. Drug reaction relationships are represented in the form of vector of weights which can be defined as a solution to some global optimization problem. This method is applied to Cardiovascular type of reactions from the Australian Adverse Drug Reaction Advisory Committee (ADRAC) database. This approach, based on the optimization approach has been implemented in software which can be used to determine the drugs that are the most likely cause of some given reactions or the reactions that are most likely to occur for a given set of drugs.

Analyzing the Structure of Biological Networks

Sergiy Butenko and Svyatoslav Trukhanov

Department of Industrial Engineering
Texas A&M University, College Station, TX

Abstract

Network models are becoming common in biology. For example, in gene networks the vertices correspond to genes and the edges represent functional relations between these genes that are identified using the comparative genomics methods. Similarly, in protein networks the vertices represent proteins and the edges connect pairs of interacting proteins.

This talk proposes new techniques for analysis of clusters in biological networks some of which are based on cohesive subgroup models in social networks. Solving the clustering problems in gene networks allows to identify groups of genes which have similar expression patterns. This information is crucial for understanding the nature of genetic diseases. Study of the structural properties of protein networks is important for understanding the cellular processes. As an example of application of the network approach, we report the results of our analysis of the network based on all known interactions in the budding yeast, *Saccharomyces Cerevisiae*, available from UCLA's Database of Interacting Proteins (DIP).

Multiscale models of NO and O2 biotransport in the microcirculation

Nikos Tsoukias

Florida International University

Abstract

The microcirculation plays a key role in the transport of oxygen from blood to tissue. The first theoretical analysis of oxygen delivery to tissue was done by August Krogh in 1919. Motivated by observations in skeletal muscle, Krogh presented a simple mathematical model for oxygen transport in capillary-perfused tissue. The Krogh model has provided many valuable insights into O₂ transport, however over the last two decades it has been shown to neglect many physiologically important aspects of microvascular O₂ delivery. In particular, it is now known that the complexity of microvascular geometry and hemodynamics, as well as blood properties (such as hemoglobin's oxygen affinity and cooperativity), can significantly affect O₂ delivery to tissue. A better fundamental understanding of the oxygen transport process is sought through the development of detailed high performance computing code of oxygen delivery by three dimensional microvascular networks. We applied this code to investigate optimal design characteristics for hemoglobin-based oxygen carriers (HBOCs). Administration HBOCs presents the most promising alternative to blood transfusion. Optimal O₂ binding properties for the transfused hemoglobin have not been determined yet. In addition, the hypertensive effects often seen after administration are considered a significant obstacle to the use of HBOCs and have been attributed to the scavenging of NO by the plasma-based Hb. Thus, future products need to address both O₂ and NO kinetics with the transfused hemoglobin molecule. Theoretical analyses and multiscale computational modeling examine the O₂/NO dynamics in the microcirculation in the presence and absence of HBOCs.

An Emerging Conceptual Framework of Cancer Gene Identification

Rongling Wu

Department of Statistics, University of Florida

Abstract

A general growth model derived from basic cellular properties can be used to describe the dynamic process of cancer growth. Here, we integrate this pervasive growth model into a statistical framework for fine mapping of quantitative trait loci underlying cancer growth based on a linkage disequilibrium-based haplotype block analysis. This integrative framework provides the basis for deciphering the genetic machinery that regulates the timing of cancer emergence, growth and differentiation.

Pattern-Match Regularity Statistics – A Measure Quantifying the Characteristics of Epileptic Seizures

Deng-Shan Shiau

Department of Neuroscience, University of Florida

Abstract

Quantitative analyses of intracranial EEG recordings from epileptic patients with temporal lobe epilepsy indicate that ictal and preictal states can be distinguished from seizure-free states for the applications of detection and prediction of seizures. These findings suggested that it is possible to develop an implantable device for diagnostic and therapeutic purposes. In this study, we present a new measure of signal regularity, pattern-match regularity score (PMRS), for the detection of EEG state changes, especially seizures. The measure is based on the estimation of signal pattern similarity. This study tests the hypothesis that PMRS can be used to distinguish state changes in intracranial EEG recordings.

Intracranial EEG recordings obtained from 6 patients with a total of 81 medically intractable partial seizures were analyzed to test the hypothesis. PMRS algorithm involves state space reconstruction, search for the pattern matched state vectors, and the estimation of pattern-match probabilities. The paired-T statistic was employed for each 10-minute sliding overlapping window to test the mean difference of PMRS values between two electrode sites. Electrode pairs were considered not entrained during any 10-minute period if the mean PMRS values were significantly different ($p < 0.05$). The PMRS and T-index curves were generated for the 1-hour time interval before and after each seizure. Significant changes observed in both PMRS and T-index curves were used for the detection of epileptic state changes in EEG recordings.

Significant decrease of PMRS values during the ictal periods was observed in 91.4% of seizures. 87.7% of the preictal periods were detected by the presence of entrainment transition (gradual decrease in T-index values), and 87.7% of the seizures showed the postictal disentrainment with a rapid increase of T-index values after the end of a seizure.

The results suggest that epileptic state changes can be detected by pattern-match regularity statistical analysis of EEG recordings from intracranial electrodes. Thus, it may be possible to predict and detect a seizure with this measure for clinical applications.

Systems biology approach to decipher neuritogenesis: ROCK pathways in mediating neurite outgrowth in PC-12 cells

Z.Q. Zhang^{1,2,3}, A. K. Ottens^{1,2,3}, S.F. Larner^{2,3}, F.H. Kobeissy^{1,2,4}, M.L. Williams³, S. Busygin⁵, A. Revenko⁵, P. M. Pardalos⁵, R.L. Hayes^{2,3,4}, K. K. W.Wang^{1,2,3,4}

Centers for Neuroproteomics and Biomarkers Research¹ and Traumatic Brain Injury Studies²; Departments of Neuroscience³ and Psychiatry⁴ of the McKnight Brain Institute; Department of Industrial and Systems Engineering⁵; University of Florida, Gainesville, FL 32611.

Abstract

Axons fail to regenerate in the adult central nervous system (CNS) following injury. A systems biology approach to decipher the signaling transduction pathways involved in neuritogenesis would help significantly in understanding the mechanisms of axonal regeneration and developing strategies to promote its regeneration in various CNS pathologies such as traumatic brain injury, stroke and Alzheimer's disease.

To achieve this goal, we utilized a novel pharmacological agent Y-27632 (Rho-associated kinase (ROCK) inhibitor), which has demonstrated the ability to promote neuritogenesis in various model systems, in the rat pheochromocytoma (PC12) cells. In this study, we observed that Y-27632 induced robust neurite outgrowth in a time- and dose-dependent manner. PC-12 cells responded to ROCK inhibition within 5 minutes at a 25 M as detected by FITC-phalloidin staining of F-actin filaments and neuronal specific 3-tubulin.

To further investigate the ROCK-related pathways in mediating neurite outgrowth, we analyzed the gene expression (genomics) and translated proteins profile (proteomics) of control and a 12hr treated with 25M concentration of Y27632 in PC-12 cells. Changes in gene transcription levels were observed by Affymetrix Microarray gene chips and data correspondence analysis. Differential protein expression analysis was accomplished by multi-dimensional separations involving biphasic ion-exchange chromatography in tandem followed by LC-MS/MS protein identification. Data comparison was performed via Contrast software. More than 40 genes and proteins, many of which are known to be associate with neuritogenesis or cell growth, were potentially found to be involved in ROCK inhibition induced neurite outgrowth in PC-12 cells. Further analysis is underway by real-time PCR and immunoblot. Through systematic application of genomic and proteomic tools, we have achieved a better understanding of neuritogenesis via the ROCK signaling pathway. Application of this prototypic network model to multi-scale networks, such as primary neurons, whole tissue, animal models and eventually clinical studies, will highly improve drug developmental efforts to promote axonal regeneration in the CNS.

Real-time selection of subset of neurons in BMIs

Sung-Phil Kim, Jose M. Carmena, Miguel A.L. Nicolelis, and Jose C. Principe.

Abstract

Recent studies in a closed-loop brain-machine interface (BMI) that use signal processing models to decode motor parameters from firing activities of multiple cortical neurons have revealed valuable results for understanding the properties of neuronal ensembles. One of them is the dynamic changes of neuronal activities related with the associated motor parameters. This property has been investigated through static statistical methods that require a block of data samples. However, more useful information can be extracted if we measure the dynamic changes on a sample-by-sample basis. An on-line variable selection rooted in linear regression enables a real-time analysis for the temporal changes of the relationship of neuronal firing activities with motor parameters. We demonstrate that this selection method can reveal how subsets of neurons are related to the specific kinetic features of arm movements.

Aligning Families of 2D-Gels by a Combined Multiresolution Forward-Inverse Transformation Approach

Xing Liu and Florian A. Potra

Abstract

A new method for aligning families of two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) images arising in proteomics studies is presented. First forward piecewise bilinear transformations are used to determine an ideal gel and to obtain an initial alignment of the family of gels to this ideal gel. Both the ideal landmarks and the coefficients defining the transformations are obtained by solving a quadratic programming problem. The alignment is then improved by using inverse transformations on finer grids. Numerical results for a family of 123 gels are reported.

Novel unsupervised methods for identifying protein coding sequence from genomic DNA sequences

Jianbo Gao, Yan Qi, Yinhe Cao, Wen-wen Tung

Abstract

Most codon indices used today are based on highly biased non-random usage of codons in coding regions. The background of a coding or non-coding DNA sequence, however, is fairly random, and can be characterized as a random fractal. When a gene finding algorithm incorporates multiple sources of information about coding regions, it becomes more successful. It is thus highly desirable to develop new and efficient codon indices by simultaneously characterizing the fractal and periodic features of a DNA sequence. In this paper, we describe a novel way of achieving this goal. The efficiency of the new codon index is evaluated by studying all of the 16 yeast chromosomes. In particular, we show that the method automatically and correctly identifies which of the three reading frames is the one that contains a gene.

Pooling Designs in Molecular Biology

Weili Wu, Yingshu Li, and Ding-Zhu Du

Abstract

The pooling design is a tool used in experiment designs of molecular biology. In this talk, we explain a new idea to construct pooling designs.

Statistical Analysis of Baroreceptors in Rabbits

Nasir Hussain

Abstract

A fundamental problem in neuroscience is how information is represented, processed, and transmitted by neurons. After detailed research and investigation, it is an established belief that neural information processing relies on the transmission of a series of stereotyped events. Basic biophysics for these action potentials (spikes) is also accepted. However the statistical features that convey this information are yet to be understood.

Research is carried out in various directions. Specific features such as the number of spikes fired by a population of neurons, their precise times of occurrence, the pattern of intervals, and various kinds of patterns of activity across a population are under extensive research. Experiments in neurophysiology return only a sample of underlying response profile. In order to get the best possible estimate of the true profile a balance is required between individual responses and collective spikes.

Bins are most commonly used balancing mechanism that is used to group responses together into category ranges of response space. Common binning strategies try to equalize bin membership, either by dividing the total response range into bins of equal size or by allowing bins of different size while trying to equalize the number of responses that fall inside each bin. In our analysis we are using the latter one.

In this research we have tried to analyze statistical parameters of spike trains from baroreceptor. Our goal was to achieve a relationship between stimulus and response and create a formulation for prediction of a spike train (response) by looking its corresponding blood pressure pulse (stimulus). We have also tried to investigate the dependence of spike in any time space (bin) to its preceding spike or spikes.

We propose a strategy for predicting stimulus response relationship in rabbits baroreceptor data. Stimulus (blood pressure pulse) and response (spikes) were collected from an adult rabbit. Each blood pressure pulse was then taken as variable window to form one trial. Spike encoding was carried out to form binary words of 33 bit length for each trial. The special occurrence and the slopes of the rising edge of pulses were taken as means of classification of variable window. Pulse windows and their corresponding spikes were then classified and trends of similarity were found in pulses and spikes of one class. Conditional probability functions were created for each class. These functions were used for class prediction of unclassified words (spike train). A comparison of prediction efficiency between conditional probability functions and firing distribution patterns was carried out.

We found that there are two parameters that affect our prediction efficiency, memory of probability functions and length of sliding window through which these functions are moved. We found that our probability functions have comparable prediction efficiency as firing distribution patterns. It was found that by using one bit memory in conditional probability functions and taking only one bit into account we got 75.6% and 54% correct prediction for Class 2 and Class 3 respectively. With dendrograms we got 75.6% and 54% percent correct results with taking only one bit into account. It was also found that taking more than two bits of memory for class prediction had only little or no benefit in correct results but our prediction was improved as compared to one bit memory. However it was observed that if the sliding window was moved from one to 12 bits the results were significantly impacted. This was also the case when window was moved through the whole thirty-three bit words. From the results of probability functions we found that occurrence of spike in any time

bin is dependent on stimulus (rising/falling) and preceding spike only. We were unable to achieve a mathematical model for spike occurrence based on found variables (stimulus and previous spike).

The Spike Activity of Neocortical Columns: A Dynamical Systems Analysis

Arunava Banerjee

Abstract

I shall begin by introducing an abstract dynamical system for networks of spiking neurons that is formulated based on a general model of the biological neuron. I shall then present simulation as well as analytical results for the class of instantiations of the system that model typical neocortical columns. Based on these results I shall argue that the spike activity of neocortical columns is profoundly influenced by attractors that are not only almost surely chaotic but are also potentially anisotropic.

*Rescaling method for quantifying the turnover rates of lymphocytes using the
CFSE assay*

Sergei S. Pilyugin, Vitaly V. Ganusov, and Rustom Antia

Abstract

The CFSE dye dilution assay is widely used to determine the number of divisions a given labeled cell has undergone both in vitro and in vivo. In this talk, I will describe how the CFSE data can be used to estimate the parameters determining cell division and death. I will also discuss the mathematical models that are commonly used to analyze the CFSE, and illustrate their advantages/shortcomings.

A column generation approach to radiation therapy treatment planning using aperture modulation

H. Edwin Romeijn, Ravindra K. Ahuja, James F. Dempsey, Arvind Kumar

Abstract

We consider the problem of radiation therapy treatment planning for cancer patients. During radiation therapy, beams of radiation pass through a patient. This radiation kills both cancerous and normal cells, so the radiation therapy must be carefully planned to deliver a clinically prescribed dose to certain targets while sparing nearby organs and tissues. Currently, a technique called intensity modulated radiation therapy (IMRT) is considered to be the most effective radiation therapy for many forms of cancer. In IMRT, the patient is irradiated from several different directions. From each direction, one or more irregularly shaped radiation beams of uniform intensity are used to deliver the treatment. This paper deals with the problem of designing a treatment plan for IMRT that determines an optimal set of such shapes (called apertures) and their corresponding intensities. This is in contrast with established two-stage approaches where, in the first phase, each radiation beam is viewed as consisting of a set of individual beamlets, each with its own intensity. A second phase is then needed to approximate and decompose the optimal intensity profile into a set of apertures with corresponding intensities. The problem is formulated as a large-scale convex programming problem, and a column generation approach to deal with its dimensionality is developed. The associated pricing problem determines, in each iteration, one or more apertures to be added to our problem. Several variants of this pricing problem are discussed, each corresponding to a particular set of constraints that the apertures must satisfy in one or more of the currently available commercial IMRT equipment. Polynomial-time algorithms for solving each of these variants of the pricing problem to optimality are presented. Finally, the effectiveness of our approach is demonstrated on clinical data.

Directions in Analysis of Large Data Sets from Synchrotron X-ray Studies of Tissue Samples

**J.F. Collingwood¹, M. Davidson², A. Mikhaylova³, J. Terry⁴, R. Barrea⁵, J. Dobson¹,
C. Batich³**

¹ Institute for Science & Technology in Medicine, Keele University, Thornburrow Drive, Hartshill, Stoke-on-Trent, ST4 7QB United Kingdom

² Department of Materials Science & Engineering, University of Florida, Gainesville, Florida USA 32611

³ Department of Biomedical Engineering, University of Florida, Gainesville, Florida, USA 32611

⁴ Biological, Chemical, and Physical Sciences, Illinois Institute of Technology, Chicago, Illinois USA 60616

⁵ Biophysics Collaborative Access Team , APS/435B/Sector 18, Argonne National Laboratory, 9700 S. Cass Avenue Argonne, Illinois, USA 60439

Abstract

Synchrotron x-rays are increasingly used to study biological systems. Whilst some fields such as molecular crystallography are established, other fields are still emerging. Our group has been creating methods to study brain metal accumulation in neurodegenerative diseases such as Alzheimer's and Huntington's disease, and the techniques have broad applicability to the study of inorganic compounds in all tissues.

The approach we have developed for tissue preparation and subsequent analysis has made it possible to go beyond the constraints of standard histochemical staining techniques, allowing identification and characterization of multiple elemental concentrations over a comparatively large area tissue section, resolving regions of interest on a micron, and hence sub-cellular, scale.

The information obtained, particularly concerning biomineral accumulations, is particularly relevant to the pathogenesis of neurodegenerative diseases, where oxidative stress is a significant factor in neuronal atrophy, and certain metals such as iron, zinc, copper and aluminium are implicated at an early stage of the disease process.

The experimental techniques employed include the collection of energy spectra at each point in a two-dimensional fluorescence map of the sample area, and the collection of diffraction patterns from a CCD area detector. This allows us to simultaneously obtain fluorescence intensity maps for a variety of elements, alongside diffraction spectra from each region. These processes generate a huge quantity of data, where each energy spectrum represents over 1MB of ASCII data (representing several GB of raw data per sample), and the 1 mega-pixel CCD camera, collecting 16 bits per pixel, records approximately 100GB of data per tissue map.

The energy spectra consist of multiple overlapping Gaussian peaks representing concentrations of the metals present in this pixel. Software is being developed that will allow simultaneous solution of the multiple linear equations at each pixel. If this analysis is performed in real time, it will be possible to reduce the data from each energy spectrum to approximately ten data points giving the intensity and relative concentrations of the elements scanned, rather than retaining the entire spectrum as a 1.24MB ASCII file. The more difficult problem of recognition of relevant diffraction spots from the CCD will require more advanced pattern recognition, and subsequent description in a more condensed format will be dealt with in a subsequent real-time data-processing program. Approximately twenty distinct diffracting compounds of interest are anticipated, primarily iron biominerals such as ferritin

cores and haemosiderin. It is therefore necessary to use a pattern recognition principal-component-type analysis rather than to solve each diffraction pattern from first principles.

Having demonstrated a combined x-ray absorption spectroscopy approach to identify a range of iron biominerals in various tissues, it is now necessary to develop software that will allow efficient processing of multiple samples. The objective is to develop automated data acquisition software that will allow real time data reduction and pattern recognition of diffraction patterns (including both diffuse rings and spots), along with reduction of the element concentration data from the energy spectra. These advances in real-time data processing will allow significant progress to be made in many areas of the biosciences without generating unfeasible volumes of data.

Single Digit Region of Interest (SDROI) Extraction by Reducing the Tissue Influence

Dong Hua, Abdou Youssef

Department of Computer Science, The George Washington University 801
22nd St. NW, Washington, DC 20052

gwuhua@gwu.edu

Abstract

Hand X-ray images are often used for measuring skeleton developmental status via bone age assessment especially in children and adolescents. It forms an evidence for diagnostic use and treatment for various diseases such as metabolic and growth abnormalities. Traditional methods include the atlas matching by Greulich and Pyle (1971) and Tanner and Whitehouse (TW2). The general impression-based atlas matching often leads to error due to subjectivity while the high complexity of TW2 limits its use to within twenty percent. Efforts have been made intensively to computerize the entire process of bone age assessment to achieve both reliability and efficiency. One significant step towards this goal is to correctly extract the region of interest (ROI), which is usually small (e.g., knuckle). However, the non-uniformity of gray distribution, noise and angled posture of hand, often make the extraction of the ROI quite complicated. One possible solution is to extract a relatively large ROI within which those small particular ROIs are further investigated. By doing so, the influences of the above-stated factors are tackled by step-by-step 'zoom-in' and hence achieve extraction of small ROIs effectively. Unfortunately, the existence of tissue, especially on the metacarpals, often results in the ineffectiveness of various techniques (e.g., edge detection). In this paper, we develop an approach to solve this issue by using spectrum analysis as well as other image processing techniques. The effectiveness is shown by experiments on single-digit-region-of-interest (SDROI) extraction. 97% accuracy is achieved on one real data set.

Simultaneous Estimation of Myocardial Intensity and Motion

B. A. Mair

Department of Mathematics
David Gilland, Nuclear & Radiological Eng.
University of Florida

Abstract

Gated cardiac emission computed tomography utilizes the detected radiation from a decaying radio-isotope to reconstruct a sequence of images of the heart wall, and the motion vector field between these images, at various stages of the cardiac cycle. The short imaging time and cardiac motion reduces the reconstructed image quality and introduces uncertainty in the estimated parameters if the images and motion are reconstructed independently of each other. In this talk we present a method that combines image reconstruction and motion estimation in a single iterative optimization algorithm. The method is based on modeling the myocardium as a deformable elastic material and combining maximum likelihood image estimation with motion estimation methods in computer vision.

Modeling and Solving String Selection Problems

Claudio N. Meneses, Panos M. Pardalos

Abstract

In this talk we present models and exact algorithms for solving two important NP-hard problems that arise in computational biology, namely:

Closest String Problem: Instance: Given a finite set $S = s^1, s^2, \dots, s^n$ of n strings of length m over an alphabet A . Objective: Find a center string x of length m over A minimizing d such that for every string s_i in S , $d_H(x, s^i) \leq d$.

Close to Most String Problem: Instance: Given a finite set $S = s^1, s^2, \dots, s^n$ of n strings of length m over an alphabet A and a threshold $k > 0$. Objective: Find a string x of length m over A maximizing the number of strings s_i in S such that $d_H(x, s^i) \leq k$.

By $d_H(x, s^i)$ we mean the Hamming distance between x and s^i . That is, given two strings s and x of same length, the Hamming distance between them is the number of mismatched positions. For example, if $s = ACT$ and $x = CCA$, then $d_H(s, x) = 2$.

Capacitance, Hardness and the Potential Energy Function: A Model for Chemical Reactivity Prediction

Alejandra Borquez¹, Cristian Cardenas-Lailhacar²

¹ Dept. of Chemical Engineering, University of Florida,
227 Chemical Engineering Bldg, PO Box 116005, Gainesville, FL 32611

² Dept. of Industrial and Systems Engineering, University of Florida
303 Weil Hall, P.O. Box 116595, Gainesville, FL 32611

Abstract

Quantum mechanics, and later on its applications to molecular systems through quantum chemistry, have used for a long time Hartree-Fock ab-initio, semi-empirical, molecular mechanics tools to study specific and sizeable systems of interest. Lately, and in particular density functional theory (DFT) has provided the theoretical basis to study reactivity parameters of molecular systems, this is, to its resistance to change its electron density distribution, namely the molecular hardness (η). Recently, Capacitance (χ) a property of molecular systems (drugs, proteins, etc.) has generated great interest because of its potential to study electronic properties of materials that can work as insulators. As in the potential energy function and hardness, its evolution along a given reaction coordinate can be examined in order to better understand the chemical reactivity of the system of interest.

In this work, compounds that are often used as prototypes to study their linkage in proteins are shown. Of particular interest are systems with sulfur - sulfur, sulfur - nitrogen, sulfur - oxygen, and nitrogen - oxygen bonds because of their importance and participation in genome (sequences), protein structure, protein-protein interaction, protein-DNA interaction, etc. Based on the DFT framework, for which the total number of electrons must be conserved, we study rotational isomerization reactions that can be regarded as a reorganization / redistribution of electron density among atoms in a system.

The relation between the molecular capacitance and hardness, with the potential energy function (E) (all global properties) is here studied. It is shown that they are all analytically related to E . Correction terms to the definition χ and η , that include a molecular symmetry parameter (b), are included. An expression for the activation capacitance and hardness, in terms of the activation potential energy function and b for the molecular system under study, are derived. A qualitative proof of the principle of maximum hardness (PMH) is shown.

Controllability of Graphs Related to the Epileptic Brain

Vitaliy A. Yatsenko

Institute of Space Research NASU and NSAU, Kiev, 03022, Ukraine

Abstract

The purpose of this report is to present an arguments and a method for controllability of neuronal networks. The novelty in this report is in the representation of the neuronal dynamics by graph models. We focus on algebraic and geometric aspects of graph controllability using Lyapunov global and local exponents and dynamical models. Also, the probability density of Lyapunov exponents (PDLE) is studied and it is shown that the PDLE gives the measure of the irregularity for the graphs. We then present the algorithms, give an effectiveness analysis of them, and describe their implementation. Finally, we give several examples and some conclusions.

Quality Assessment in Feature Selection

Cheong Hee Park, Moongu Jeon, Panos Pardalos, Haesun Park

Abstract

Feature selection is a task to select a subset of features which is the most essential for intended tasks such as classification, clustering or regression analysis. In gene expression microarray data, being able to select a few of genes not only makes data analysis efficient but also helps biological interpretation. However, microarray data has typically several thousands of genes (features) but only tens of samples, and problems which can occur due to small sample size have not been addressed well. We discuss some issues on feature selection in microarray data: whether choosing optimal number of features and ranking genes is reliable and meaningful when different feature selection methods and different setting of training and test sets are used.

Quadratic-type models and upper bounds for the problems of finding the maximum weighted independent set in graphs

N.Z. Shor, P.I. Stetsyuk

V.M. Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine, Kyiv

Abstract

Let a vertex weighted undirected graph $G_w(V, E)$ be given: V is the set of vertices, E is the set of edges, $w = \{w_i\}_{i \in V}$ are the positive weights of vertices. The maximum weighted independent (stable) set problem in graph $G_w(V, E)$ can be reduced to a multiextremal quadratic-type nonlinear programming problem:

$$\begin{aligned} \alpha(G_w) &= \max \sum_{i \in V} w_i x_i; \\ \text{s.t.} \quad & x_i x_j = 0; \forall (i, j) \in E; \\ & x_i^2 - x_i = 0; \forall i \in V. \end{aligned} \tag{P1}$$

To solve the problem (P1) one may use Lagrangian bounds technique in combination with branch and bounds method. The problems of finding Lagrangian bounds (upper bounds for $\alpha(G_w)$) can be reduced to minimization of nonsmooth convex functions and can be solved by methods of non-differentiable optimization. Lagrangian bounds may be improved by adding to the problem (P1) so-called superfluous quadratic constraints, which modify Lagrangian bounds. For example, we can use superfluous quadratic inequalities of the following form:

$$x_i x_j = 0 \text{ for all (some) nonadjacent pairs } (i; j)$$

or

$$x_k(x_i + x_j) \leq x_k; \text{ for any } k \in V \text{ and } (i; j) \in E.$$

The use of the standard method of finding Lagrangian bounds for some of these quadratic-type formulations of the problem of finding the maximum weighted independent set in a graph gives the same results as obtained by the orthonormal representation of graphs.

It is illustrated how Lagrangian approach in combination with superfluous quadratic constraints improve upper bounds for $\alpha(G_w)$. Analysis of numerical experiments is given.