Combinatorial Optimization in Computational Biology

Dan Gusfield

Computer science department, University of California at Davis, USA

Abstract: Combinatorial Optimization is a central sub-area in Operations Research that has found many applications in computational biology. In this talk I will survey some of my research in computational biology that uses graph theory, matroid theory, and integer linear programming. The biological applications come from haplotyping, the study of recombination and recombination networks, and phylogenetics.

Network-Based Systems Biology

Luonan Chen

1. Institute of Systems Biology, Shanghai University, Shanghai 200444, China;
2. Department of Electrical and Electronics Engineering, Osaka Sangyo University, Osaka 574-8530, Japan;
3. ERATO Aihara Complexity Modelling Project, JST, Tokyo 151-0064, Japan
4. Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan

Abstract: Burgeoning high-throughput data are driving the integrative study from describing phenomena to understanding design principle and essential mechanism, from studying individual components to understanding functional networks for biological systems, cells, organs, and even organisms. Network-based systems biology is an emerging area focusing on various biomolecular networks, which is also a multidiscipline intersection of mathematics, computer science and biology. To elucidate the fundamental mechanisms of cellular systems, study on biomolecular networks is increasingly attracting much attention from many academic communities, such as mathematics, information science and life science. A major challenge in network-based biology or simply network biology is to investigate how cellular systems facilitate biological functions and achieve complex lives by various interactions (pathways and networks) between genes, proteins and metabolities at network level. By developing theoretical and computational methodologies, network-based systems biology or simply network biology is to study an organism, viewed as a dynamical or interacting network of genes, proteins and biochemical reactions, which eventually give rise to life. In contrast to the study on individual molecules, biomolecular networks governed by universal laws offer a new conceptual framework that could potentially revolutionize our view of biology and pathologies. Therefore, it is highly demanded that mathematicians and computer scientists provide theoretical and computational methodologies to reveal the essential biological mechanisms of living organisms from a system or network perspective.

   I will demonstrate our recent works on the area of network biology, including topics on (1). modeling and analyzing biomolecular networks, (2). controling biological networks, (3). synchronizing biological networks, (4). quantifying   biological networks, (5). reverse-engineering biomolecular networks, (6). forward-engineering biomolecular networks, (7). aligning biomolecular networks, (8). identifying functional subnetworks, (9). network-based functional analysis, (10). network-based drug target detection; in particular, stressing on the aspects of network, system, integration and engineering at molecular level. The related papers can be found from http://www.isb.shu.edu.cn , http://intelligent.eic.osaka-sandai.ac.jp/ or http://zhangroup.aporc.org/   .

References

[1]. L.Chen, R-S.Wang, X-S.Zhang. *Biomolecular Network*: Computational Methods and Applications in Bioinformatics and Systems Biology, *John Wiley & Sons*, 2009.

[2]. L.Chen, R-Q.Wang, C.Li, K.Aihara. *Modelling Biomolecular Networks*: Structures and Dynamics. ***Springer-Verlag***, London, 2009.

Neurodynamic Optimization: New Models and Selected Applications

Jun Wang

Department of Mechanical & Automation Engineering
The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

Abstract: Optimization problems arise in a wide variety of scientific and engineering applications. It is computationally challenging when optimization procedures have to be performed in real time to optimize the performance of dynamical systems. For such applications, classical optimization techniques may not be competent due to the problem dimensionality and stringent requirement on computational time. One very promising approach to dynamic optimization is to apply artificial neural networks. Because of the inherent nature of parallel and distributed information processing in neural networks, the convergence rate of the solution process is not decreasing as the size of the problem increases. Neural networks can be implemented physically in designated hardware such as ASICs where optimization is carried out in a truly parallel and distributed manner. This feature is particularly desirable for dynamic optimization in decentralized decision-making situations. In this talk, we will present the historic review and the state of the art of neurodynamic optimization models and selected applications. Specifically, starting from the motivation of neurodynamic optimization, we will review various recurrent neural network models for optimization. Theoretical results about the stability and optimality of the neurodynamic optimization models will be given along with illustrative examples and simulation results. It will be shown that many computational problems, such as k winner-take-all, can be readily solved by using the neurodynamic optimization models.

Time-series-based Ensemble Modeling for Bio-Medical Applications

Maciej Ogorzalek, Christian Merkwirth, and Joerg Wichard

AGH University of Science and Technology, Poland

Abstract: We propose toes use ensembles of models constructed using methods of Statistical Learning. The input data for model construction consists of real measurements taken in physical system under consideration. Further we propose a program toolbox which allows the construction of single models as well as heterogenous ensembles of linear and nonlinear models types. Several well performing model types, among which are ridge regression, k-nearest neighbor models and neural networks have been implemented. Ensembles of heterogenous models typically yield a better generalization performance than homogenous ensembles. Additionally given are methods for model validation and assessment as well as adaptor classes performing transparent feature selection or random subspace training on large number of input variables. The toolbox is implemented in Matlab and C++ and available under the GPL. Several applications of the described methods and the numerical toolbox itself are described. These include ECG modeling, classification of activity in drug design and image-based melanoma diagnosis.

Establishment of a HCV biological database for epitope-paratope matching using phage display technology

Te-Fen Kao, **Hao-Teng Chang**

Graduate Institute of Molecular Systems Biomedicine, Department of Biological Science and Technology, China Medical University, Taichung, Taiwan

Dr. Hao-Teng Chang: htchang@mail.cmu.edu.tw

Abstract: Hepatitis C virus (HCV) is Hepacivirus genus within the Flaviviridae families. It is known that at present, 6 genotypes are determined. Ten viral proteins including core, E1, E2, p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B proteins are translated to one polypeptide and then cleaved by NS3 protease. One additional special protein, called alternate reading frame protein (ARFP) or F protein, is produced from +1 frame shift in core protein and its function is unclear. In this study, NS2 and F protein were analyzed by bioinformatics software, MISA (Multiple Indexing Sequence Alignment) and LEPD (Liner Epitope Prediction Database) and 3 peptides, 2 for F protein and 1 for NS2, with high antigenicity and consensus between 6 genotypes were selected as antigenic peptides for paratope screening. Phage display is a powerful tool to define protein-protein interactions by generating peptide binders against target antigens. It was employed to pick out the sequences of variable regions (Fv) of human antibodies that can interact with the designed antigenic peptides. Finally, the sequences of F- and NS2-interacting paratopes were identified and the biological database which correlated the F or NS2 with paratopes was generated to further bioinformatics studies.

Keywords: epitope, paratope, phage display

Phylogentic networks: Methods to compute split systems

Stefan Grünewald

CAS-MPG Partner Institute for Computational Biology, Shanghai, China
stefan@picb.ac.cn

Abstract: Phylogenetic networks are a generalization of phylogenetic trees that allow the representation of conflicting signals or alternative evolutionary histories in a single diagram. Networks are commonly used if the underlying history is not treelike. For example, recombination, hybridization, and gene transfer can all lead to histories that are not adequately represented by a single tree. Moreover, even when the history is treelike, parallel evolution, model heterogeneity, and sampling error can make it hard to find a unique tree. In such cases networks provide a tool for representing ambiguity or for visualizing a collection of feasible trees. In the unrooted case, splits graphs are the most popular class of phylogenetic networks. I will introduce several methods to construct splits graphs from various kinds of input data. In particular, I will present the quartet-based QNet algorithm as well as SuperQ, a method that constructs supernetworks, i.e. it puzzles together trees or networks with overlapping taxa sets into one parental network.

An Integer Linear Programming Approach to Topological Alignments and its Applications

Axel Mosig

Pattern Discovery in Biology Group, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, 200031 Shanghai, P.R. China

Abstract: We introduce a class of so-called topological alignment problems and their application to tracking cells in bioimaging video sequences. In their simplest form, topological alignments read as a generalized bipartite matching problem: rather than finding edges between two vertices, topological alignments find edges between sets of vertices. Our computational approach towards solving these matching problems is based on an integer linear programming formulation, which can be easily adopted to more general scenarios: if the vertices are derived from cycles in a planar graph, they lead to finding largest common planar subgraphs, or, by introducing further constraints, even to largest common planar subembeddings. While such problems are computationally hard in general, we have identified instances that are tractable in practice for our cell tracking algorithm.