

BOOK REVIEW

Systemic Approaches in Bioinformatics and Computational Systems Biology: Recent Advances

*Reviewed by Christos Makris, Department of Computer Engineering and Informatics,
University of Patras, Patras, Greece*

*Systemic Approaches in Bioinformatics and
Computational Systems Biology: Recent
Advances*

*Paola Lecca, Dan Tulpan, and
Kanagasabai Rejaraman*

© 2011 by IGI-Global

264 pp.

\$265.00

ISBN 978-161-3504-35-2

INTRODUCTION

Bioinformatics is an interdisciplinary field that combines Molecular Biology and Computer Science, and focuses in the management of the information contained in biological macromolecules (DNA molecules, genes and proteins) using computational techniques from the areas of data structures, machine learning, data bases, information retrieval and computational geometry. Computational Systems

Biology, deals with modeling complex interactions between biological systems, employing a holistic approach, in order to understand how the various properties of a biological system as a whole, emerge. This specific volume is comprehensive in quality and quantity study concerning systemic approaches in the areas of Bioinformatics and Computational Systems Biology. The volume contains a series of articles that explain algorithmic approaches that emerge when developing systemic approaches; the articles cover a broad range of topics that could benefit both theoretical insight and practical applications in the Bioinformatics' community.

ORGANIZATION OF THE BOOK

In particular the book consists of eleven chapters each exploiting a different theme of the specific research area.

The first chapter is written by Miroslava Cuperlovic-Culf and discusses unsupervised data analysis methods employed in Metabolomics and Metabonomics. A survey concerning the pros and cons of various clustering techniques and data analysis methods such as PCA (Principal Component Analysis), HCL (Hierarchical Clustering), KM (k-means), SOM (Self Organized Maps) and FKM (Fuzzy k-means) is presented. The survey is both theoretical and practical, with experiments on a large set of carefully produced synthetic and real data, and interesting conclusions concerning the functionality of the various algorithms are assembled. The main conclusion drawn is the usefulness of PCA when combined with other techniques, the superiority of bagged k-means in quantified data when compared to k-means, the excellence in performance of SOM, and the high efficiency of Fuzzy k-means; these three methods should be explored in metabolic applications.

The second chapter, written by Jose Garcia-Manteiga provides an overall overview of the area of data analysis in metabolomic studies. It can be considered a logical continuation of the previous chapter with a special emphasis in the data collection task. In particular it deals with study design, sampling and quenching methods, extraction procedures, and it describes in depth analytical techniques such as NMR, Mass Spectrometry, data processing and analysis. Then it concentrates on statistical analysis techniques such as data pre-processing, data analysis, PCA, Hierarchical Cluster Analysis, supervised analysis, machine learning tools, decision trees, Artificial Neural Networks (ANN) and Support Vector Machines. Moreover it describes the databases and mining software solutions for data interpretation with special focus on the modeling uses of metabolomics data for systems biology and the integrative approaches with other disciplines such as proteomics. Finally it presents future directions dealing with a combination of NMR and Mass Spectrometry, integrative multi-omics analysis, and single cell metabolomics. The main conclusion is that in

order to integrate metabolonomics and other systemic approaches a large number of scientific disciplines has to be combined.

The third chapter written by Dan Tulpan, Athos Ghiggi and Roberto Montemanni deals with computational sequence design techniques for DNA Microarray Technologies. The article presents in an understandable and comprehensive way the basic algorithms in microarray probe design, it evaluates existing probe sequences used in commercial arrays and suggests improved methodologies. From the combinatorial approaches discussed special emphasis is given at the Stochastic Local Search method, and at a Seed Building procedure. Both extensions of them to thermodynamic constraints are tested on Affymetrix microarrays. The experimental results are encouraging and depict that significant improvements can be attained if constraints based on pairwise sequence hybridization are used during the probe design process.

The fourth chapter written by Boris Jankovic, John Archer, Rajesh Chowdhary and Ulf Schaefer deals with promoter structures conserved between Homo Sapiens, Mus Musculus and Drosophila melanogaster. The authors depict the existence of shared families of DNA sequence motifs and their further study leads to various interesting outcomes such as the identification of several transcription factor binding sites and the provision of possible links to other genes for a number of fly genes that are not annotated yet. In particular their analysis identified "significant sequence conservations among the promoter regions of orthologous human and mouse gene pairs and promoter regions within the fly genome expressed in terms of shared combinations of MFs".

The fifth chapter written by Haitham Ashoor, Rajesh Chowdhary, Arturo Mora, John Archer, Karim Awara, Vladimir Bajic and Boris Jankovic deals with the identification of Translation Initiation Sites (TISs) that delimit the borders of protein coding regions; in particular they concentrate on TIS recognition directly from genomic data. They develop an

algorithmic approach that is quite accurate (in *Arabidopsis thaliana* genome it has an accuracy of over 83 percent, and in *A. thaliana* data has an accuracy of 91 percent). The authors select a subset of the originally considered features, utilizing Genetic Algorithms and employ an Artificial Neural Network-based binary classifier using a selected subset of the originally considered features. Experimental results depict that the model conserves key features and it provides the best accuracy results in the literature for the corresponding genomic sequence, thus providing a base on which to build more general models.

In the sixth chapter written by Paola Lecca and Alida Palmisano the authors build on the premise that the analysis of the network structure describing the interaction of biological entities and its evolution in time is vital in order to understand the living systems' characteristics. The article describes network inference goals and methodological approaches (transcriptional regulatory networks, signal transduction networks, metabolite networks, intraspecies or interspecies communication networks, protein-protein interaction networks), putting special emphasis on protein-protein interaction networks and building on methods such as perturbation methods, correlation-based methods, and bayesian methods. Moreover and concerning the various calibration parameters the authors report their theoretical models on parameter inference, as it is targeted on the theoretical framework of the software KInfer. The authors conclude that "Network inference is the link between the *in vivo* and *in vitro* systems biology and the *in silico* systems biology."

In the seventh chapter written by Ferenc Jordán, Carmen Maria Livi, Paola Lecca the theme of structural and dynamical heterogeneity, as it appears in ecological networks is studied. In complex ecological systems, which are composed of several components and multiple parallel interactions among them, it is increasingly needed to precisely understand structural and dynamical variability among components. This variability is the basis of adaptability and evolvability in nature, as well as adaptive

management-based applications. The authors discuss how to quantify and characterize the structural and dynamical variability in ecological networks.

The eighth chapter written by Marco Scotti deals with the presentation of the stochastic simulations to extend food web analyses. The main combinatorial structures presented is the dominator tree, a structure that is useful in presenting dominance relationships in a hierarchical digraph, while stochastic simulations based on a process algebra-based method are introduced. A case study on a real food web is performed to test the various algorithms emphasizing that "Ecosystem management would benefit from novel computational tools that allow researchers to extend stochastic-based dynamics towards spatial and temporal simulations".

In the ninth chapter written by Ozan Kahramanoğlu the author deals with process algebras and their suitability in modeling biology processes with a case study the case of phagocytosis, where cells engulf larger particles. The authors claim that process algebras can have a broader usefulness in the area of systems biology and provide a proof of concept for their claim by providing example models; they also point out the deficiencies of the approach that should be handled in order to bring the process algebra to a higher level.

The tenth chapter written by Parthasarathy Subhasini, Bernadetta Kwintiana Ane, Dieter Roller, and Marimuthu Krishnaveni deals with machine learning techniques and in particular with fusion of intelligent classifiers and their applicability in enhancing recognition of genes and protein pattern of hereditary diseases. The authors provide a nicely written survey of the various notions concerning classification algorithms, their assembling (bagging and boosting) and they develop the classifier fusion model (combination of soft-output classifiers using methods like majority vote, decision template, Bayes, and Algebraic combiners), that is trained through supervised learning techniques like multilayer perceptron (MLP) and support vector machine (SVM), in order to recognize genes

and in producing protein patterns of hereditary diseases. The application of these classifiers provides good and prospective results.

Finally in the eleventh chapter written by Yue Wang Webster, Ernst Dow and Mathew Palakal the authors deal with translational research and summarize three aspects of it: use cases/opportunities, data types and challenges, available tools and technologies. They propose Complex System as the basis of such foundation and with that they present the design decisions of a multilayer architecture called HyGen (Hypotheses Generation Framework). The authors evaluate the framework by performing tests in the Colorectal Cancer disease area and the results demonstrate the potential of this hybrid approach to reveal hidden connections and provide new drugs, biomarkers and therapies. The chapter provides a survey concerning Data sources involved in translational research,

semantic web, graph analysis, general information connecting information silos and people silos, and solutions and recommendations. Moreover it describes future research directions concerning Web2.0 and emerging sources for translational research, agent-based information extraction, edge properties and profile-driven ranking criteria.

SUMMARY

In conclusion the volume covers a broad range of topics in a well structured way, while the material both theoretical and practical is being presented in an appropriate academic way, and contains novel content useful to both theoreticians and to the developers of practical applications.

Christos Makris is an Assistant Professor in the Department of Computer Engineering and Informatics, School of Engineering, University of Patras, Greece. His research interests include Data Structures, Bioinformatics, Web Algorithmics, Computational Geometry, Data Bases and Information Retrieval. He has published over 90 papers in various scientific journals and refereed conferences and has over 350 references.