

Guest Editorial

Special Issue on Advances in Bioinformatics, Biomedicine, and Health Informatics

THIS SPECIAL issue of the IEEE TRANSACTIONS ON NANOBIOSCIENCE comprises 10 papers that are extended versions of accepted papers selected from the 2011 IEEE International Conference on Bioinformatics & Biomedicine (BIBM). BIBM 2011 received 299 research paper submissions from 36 countries and it provided a leading forum for disseminating the latest research in bioinformatics, biomedicine and health informatics. The Program Committee from 19 countries accepted 58 regular research papers in total. Since 2007, BIBM has established itself as the premier and top tier multidisciplinary research conference that brings together academic and industrial scientists from computer science, biology, medicine, mathematics, and statistics. In BIBM, open issues are addressed and research results are exchanged in all aspects of bioinformatics, biomedicine, and health informatics. BIBM provides a well-known forum for the presentation of work in databases, algorithms, interfaces, visualization, modeling, simulation, ontology, and other computational methods, with emphasis on applications in high throughput data-rich areas in biology, biomedical engineering, and health informatics.

This special issue discusses the most challenging issues in Bioinformatics including sequence data analysis, classification and machine learning, network analysis, and software.

The first group of papers is about sequence analysis. The first paper by Ozcaglar *et al.* studies how biomarkers of Mycobacterium tuberculosis complex (MTBC) mutate over time. Among the biomarkers of MTBC, spacer oligonucleotide type (spoligotype) and Mycobacterium Interspersed Repetitive Unit (MIRU) patterns are commonly used to genotype clinical MTBC strains. They presented an evolution model of spoligotype rearrangements using MIRU patterns to disambiguate the ancestors of spoligotypes. The second paper by Chen *et al.* uses the probabilistic topic models to estimate functional groups in human gut microbiome. The probabilistic topic modeling is a Bayesian method that is able to extract useful topical information from unlabeled data. Estimating the probabilistic topic model can uncover the configuration of functional groups in each sample.

The second group of papers is about classification and machine learning problems, including classification and prediction. The first work by Wu *et al.* proposes a stratified sampling method for feature subspace selection to generate decision trees in a random forest for high dimensional genome-wide association (GWA) case-control data of complex disease. The key idea was to design an equal-width discretization scheme for informativeness to divide SNPs into multiple groups. The advantage of this stratified sampling procedure can make sure each

subspace contains enough useful SNPs, but can avoid a very high computational cost of exhaustive search, and maintain the randomness of a random forest. The second paper by Liu *et al.* proposes an efficient algorithm, Combined Rule Extraction and Feature Elimination (CRF), based on 1-norm regularized random forests. CRF simultaneously extracts a small number of rules generated by random forests and selects important features. CRF was shown to select and interpret the most important information from multiple biological sources. The third paper by Li *et al.* focuses on prediction of protein subcellular localization. In order to predict the location of a protein within a cell, they formulated prediction of subcellular localization of multiplex proteins as a multi-label learning problem and then showed the methods exploiting label correlations performs better than those leveraging label-specific features.

The third group of papers focuses on the network analysis. The paper by Wang and Zheng investigates the existence of date and party hubs in protein-protein interaction networks using the information extracted from known three-dimensional structures of protein interactions. They examined the correlation of genomic essentiality, gene coexpression, and functional semantic similarity with date and party hubs and provided more conclusive information regarding date and party hubs in protein-protein interaction networks. The paper by Tian *et al.* demonstrates that stability is essential for designing or controlling genetic regulatory networks by using nonlinear differential equations. Especially they focus on delay-dependent stability of genetic regulatory networks and presented some delay-dependent stability conditions for genetic regulatory networks based on linear matrix inequality (LMI) approach. Finally, the paper by Kim and Gelenbe presents a new reverse engineering approach based on Bayesian model averaging technique which ensembles all appropriate models describing interactions among genes. The proposed Bayesian approach with a prior based on Gibbs distribution was shown to provide an efficient way to integrate multiple sources of biological data.

The final group of papers introduces new software. BioVLAB-MMIA is for a cloud environment for microRNA and mRNA integrated analysis (MMIA) on Amazon EC2. Lee and Kim extend the BioVLAB cloud workbench to develop an environment for the integrated analysis of microRNA and mRNA expression data, named BioVLAB-MMIA. The integration of microRNA expression data sets with gene expression profiles is expected to be a key research problem in life science research. ENteric Immunity SIMulator (ENISI), developed by Wendelsdorf *et al.*, is a new software platform for simulating gastroenteric infections immune mechanisms in response to resident commensal bacteria as well as invading pathogens and the effect on the development of intestinal lesions. Both

BioVLAB-MMIA and ENISI are expected to be widely used in many bioinformatics and biomedicine researches.

These ten accepted papers in this volume provide scientists with an overview on the recent advancements in bioinformatics and biomedicine. Each paper was reviewed by at least five experts in the field. We wish to thank all the reviewers for their distinguished reviewing efforts. We hope the papers can encourage researchers a more extensive use of these bioinformatics techniques and software for research in biology and medical sciences.

List of papers

Classification

== Sequence analysis ==

- 1 “Inferred Spoligoforest Topology Unravels Spatially Bimodal Distribution of Mutations in the DR Region”
- 2 “Estimating Functional Groups in Human Gut Microbiome With Probabilistic Topic Models”

== Classification/Learning ==

- 3 “SNP Selection and Classification of Genome-Wide SNP Data Using Stratified Sampling Random Forests”
- 4 “Combined Rule Extraction and Feature Elimination in Supervised Classification”
- 5 “Multilabel Learning for Protein Subcellular Location Prediction”

== Network analysis ==

- 6 “Correlation of Genomic Features With Dynamic Modularity in the Yeast Interactome: A View From the Structural Perspective”

- 7 “Robust and Global Delay-Dependent Stability for Genetic Regulatory Networks With Parameter Uncertainties”
 - 8 “Reconstruction of Large-Scale Gene Regulatory Networks Using Bayesian Model Averaging”
- == Software Application ==
- 9 “BioVLAB-MMIA: A Cloud Environment for microRNA and mRNA Integrated Analysis (MMIA) on Amazon EC2”+ mRNA)
 - 10 “ENteric Immunity Simulator: A Tool for *In Silico* Study of Gastroenteric Infections”

JIMMY XIANGJI HUANG, *Guest Editor*
 School of Information Technology
 York University
 Toronto, ONM3J 1P3 Canada

TAESUNG PARK, *Guest Editor*
 Department of Statistics
 Seoul National University
 Seoul 151-747, Korea
 National Creative Research Lab for Bioinformatics
 and Biostatics
 Seoul National University
 Seoul 151-747, Korea