Data mining in bioinformatics, biomedicine, and healthcare informatics

Guest Editorial: Data Mining in Bioinformatics, Biomedicine, and Healthcare Informatics

BIOINFORMATICS and biomedical research works are fundamental to our understanding of complex biological systems, impacting the science and technology of fields ranging from basic science to pharmaceutical and clinical sciences. This type of research requires close collaboration among multidisciplinary teams of researchers in physics, computer science, statistics, engineering, life sciences and medical sciences, and their interfaces. In this special issue on *Data Mining in Bioinformatics, Biomedicine, and Healthcare Informatics*, four papers in their significantly extended versions were selected from the papers presented at the 2012 IEEE Conference on Bioinformatics and Biomedicine and the other four came from open solicitation with a wide range of authors. These papers present original research work in bioinformatics and computational systems biology to make use of large-scale biological data or high throughput data to understand the mechanism of biological systems and to show their usefulness in diagnosis and drug design for complex diseases.

Computational modeling has played a critical role in biomedical research at different levels, from basic molecules dynamic behavior, cell–cell interaction, cell proliferation, signal transduction, to tissue and organ development. T. Peng, H. Peng, D. S. Choi, J. Su, C.-C. (Jeff) Chang, and X. Zhou investigated how myeloma initiating cells drive tumor growth, differentiation, metastasis, and recurrence of the diseases. A new lineage model was developed to model the cell–cell interactions by considering secreted factors, self-renew, and differentiation among different cell types.

S. Li, J. Nyagilo, D. Dave, and J. Gao provided a complete and deep analysis of different modeling methods for the quantitative analysis of nanoparticle surface-enhanced Raman spectra for protein biomarker detection. The fundamental differences and connections of commonly used modeling methods when applied to Raman signals were investigated with the conclusion that partial least-squares regression is the best choice for Raman spectra problems. As a different application, D. He and D. Parker investigated how different biomedical topics influence each other in the biomedical literature. A SemInf model was developed using burst-based action to model popularity and influence with an efficient solution being followed for identifying influential topics.

Genetic association analysis of complex diseases reveals the fundamental correlations at the molecular and clinical syndrome levels. However such association studies have been limited by heterogeneity in the clinical manifestations and genetic etiology. J. Sun, J. Bi, and H.R. Kranzler presented a multiview comodeling data analytic method that integrates clinical features and genetic markers of the same patient which will improve disease subtypes discovery.

Toward the final goal of providing clinical and translational supports for patients, this special issue contains three papers centered around this. Z. Huang, M.-L. Shyu, J. M. Tien, M. M. Vigoda, and D. J. Birnbach demonstrated a knowledge-assisted sequential pattern analysis system with heuristic parameter tuning to predict the changes in intrauterine pressure, which hopefully will lead to labor contraction prediction. X. Cai, J. Wei, G. Wen, and Z. Yu presented a local and global preserving semisupervised dimensionality reduction based on random subspace method for cancer classification using microarray datasets.

As an another translational application of bioinformatics tool, J. Reps, J. M. Garibaldi, U. Aickelin, D. Soria, J. E. Gibson, and R. B. Hubbard showed how a computational metaanalysis framework can signal many known rare and serious side effects for the selection of drugs investigated. The system integrated existing methods, knowledge from the web, metric learning, and semisupervised clustering.

To improve the coverage and quality of existing background knowledge base in electronic medical records, S. Perera, C. Henson, K. Thirunarayan, A. Sheth, and S. Nairy have developed a semantics driven semiautomatic method which can identify the absence of causal relationships between symptoms and disorders in background knowledge. This methodology will rectify missing relationships and will minimize the burden on domain experts.

The Guest Editors and the Editor-in-Chief would like to thank all authors for the high-quality work they contributed to this special issue, the managing editors for handling the entire paper reviewing processing, and all anonymous reviewers for their great efforts and expert comments in evaluating the papers.

JEAN X. GAO, *Guest Editor*
University of Texas at Arlington
Arlington, TX 76019 USA
gao@uta.edu

WERNER DUBITZKY, *Guest Editor*
University of Ulster
Londonderry BT48 7JL, United Kingdom
W.Dubitzky@ulster.ac.uk

Digital Object Identifier 10.1109/JBHI.2014.2306988