
Editorial

Sun Kim

School of Informatics and Computing,
Indiana University,
Bloomington, Indiana, USA
E-mail: sunkim2@indiana.edu

Biographical notes: Sun Kim is an Associate Professor and Director of Center for Bioinformatics Research in the School of Informatics and Computing at Indiana University, Bloomington, USA. He received his BS and MS and PhD in Computer Science from Seoul National University, Korea Advanced Institute of Science and Technology (KAIST) and the University of Iowa, respectively. His research interests are development and applications of machine learning algorithms to the analysis of microbial genome and cancer epigenomics data and also on the bioinformatics systems architecture in the cloud computing environments.

The rapid accumulation of biological data requires more sophisticated, efficient computational methods at almost all stages of research in biology and medical sciences from handling and analysis of data to mining novel insights on new knowledge in biology. This special edition of the *International Journal of Data Mining and Bioinformatics* contains six papers that report how bioinformatics techniques can be used for various aspects of biological research. The six papers were selected and extended among top papers from 233 submissions at the third IEEE International Conference on Bioinformatic and Biomedicine in 2009.

The first paper by S. Lee, E. Lee, K.H. Lee and D. Lee, 'Predicting disease phenotypes based on the molecular networks with Condition-Responsive Correlation', used network analysis methods for the integrated analysis of molecular interaction networks and gene expression profiles. The proposed method constructs a network where nodes denote discriminative expression levels and edges represent condition-responsive correlations between two phenotypes. The proposed network-based approach has shown improved performances over conventional gene-centric classification methods for predicting disease phenotypes.

The second paper by S. Pounds, X. Cao, C. Cheng, J. Yang, D. Campana, C-H. Pui, W.E. Evans and M.V. Relling, 'Integrated analysis of pharmacologic, clinical and SNP microarray data using Projection Onto the Most Interesting Statistical Evidence with Adaptive Permutation Testing', extended a recent method called Projection Onto the Most Interesting Statistical Evidence (PROMISE) to perform the integrated analysis of pharmacologic, clinical and genome-wide genotype data that incorporate knowledge about the biological relationships among pharmacologic and clinical response data. The proposed method was applied to a paediatric leukaemia data set and identified SNPs that have a biologically meaningful pattern of association with multiple endpoint variables.

The third paper by Q. He, K. Karsch and Y. Duan, 'Semi-automatic 3D segmentation of brain structures from MRI', presented a semi-automatic three-dimensional segmentation method for brain structures from Magnetic Resonance Imaging (MRI). The proposed method used a novel two-phase approach combining boundary-based and region-based methods and achieved more efficient segmentation. The segmentation method used a probability map effective for the brain structures with low-intensity contrast to the background and it also allows manual adjustment for segmentation so that experts can make final decisions.

The fourth paper by Y. Lu and J.L. Ram, 'Predictions of flexible C-terminal tethers of bacterial proteins with the FLEXTAIL bioinformatics pipeline', used the FLEXTAIL bioinformatics pipeline for the analysis of bacterial chemotaxis receptors to predict flexible tethers and their binding motifs based on the properties of flexibility and sequence conservation. In four groups of bacterial genomes, the algorithm identified more than 100 putative binding domains, including known bacterial chemotaxis receptor NWETF binding motifs and also some potential C-terminal flexible regions that have not previously been recognised to function as protein tethers.

The fifth paper by L. Lan and S. Vucetic, 'Improving accuracy of microarray classification by a simple multi-task feature selection filter', proposed a multitask feature selection filter that borrows strength from auxiliary microarray data from the public domain for the analysis of a small-sample microarray data. It used Kruskal-Wallis test on auxiliary data and ranked genes based on their aggregated p -values. The top-ranked genes are selected as features for the target task classifier. The multi-task filter was evaluated on microarray data related to nine different types of cancers.

The final paper by T. Fayruzov, J. Janssen, D. Vermeir, C. Cornelis and M. De Cock, 'Modelling gene and protein regulatory networks with Answer Set Programming', proposed a novel Answer Set Programming based approach to the regulatory network discovery problem. The paper showed that the proposed method can mimic Boolean network behaviour and apply it to model the Budding Yeast and Fission Yeast cell cycle networks. Final steady states of these networks were shown to correspond to those of the Boolean networks.

The six papers in this volume provide scientists with a glimpse on how bioinformatics techniques are used for research in biology and medical sciences, and I hope the papers can boost a more extensive use of bioinformatics techniques for research in biology and medical sciences.