SPECIAL ISSUE ON DATA MINING AND PATTERN ANALYSIS IN COMPUTATIONAL BIOSCIENCE

Computational bioscience aims to develop innovative methods for solving significant problems in the biological sciences. Some of the grand challenges in computational bioscience include mining molecular interactions, whole genome comparison, sequence and structure motif discovery, and gene expression microarray data analysis.

This special issue provides a collection of papers that report recent advances in computational bioscience with a focus on biological pattern discovery and data mining.

The special issue begins with a meeting report, followed by five articles. In "The DNA–Proteome: Recent advances towards establishing the protein–DNA interaction space," Erich Grotewold and Herbert Auer summarize findings from the DNA–Proteome Barcelona BioMed Conference held in Barcelona in April 2009. This conference brought together 150 scientists from around the world, exploring advances in establishing the DNA–protein space in eukaryotic organisms, from humans to yeast and plants.

Then the special issue contains two papers, both of which are related to molecular interactions. In "Prediction of protein function from connectivity of protein interaction networks," Lei Shi *et al.* present an artificial neural network approach to predict protein functions through integration of several protein interaction data sets. The authors experimentally show that their approach outperforms other existing methods on MIPS functional categories. In "Comparison of chemical descriptors for protein–compound interaction prediction," Jintao Zhang and Jun Huan report a case study on comparing the performance of several different chemical descriptors for predicting protein–compound interactions. The authors conclude that frequent subgraph-based descriptors and the signature molecular descriptor work well, and the appropriate selection of chemical descriptors is important in achieving good results.

The next two papers are concerned with pattern search in genomic data. In "Mining roX1 RNA in *Drosophila* genomes using covariance models," Kevin Byron *et al.* propose a methodology for finding roX1 non-coding RNAs in 12 *Drosophila* species by utilizing structural alignment and statistical profiles. The authors experimentally show that their methodology is more effective than Blast in detecting functional RNA homologs in *Drosophila* genomes. In "An integrated bioinformatics approach to the discovery of cis-regulatory elements involved in plant gravitropic signal transduction," Xiaoyu Liang *et al.* describe techniques for identifying putative regulatory functional elements, including transcription factor binding sites and cis-regulatory modules involved in gravitropic signal transduction. By analysing gene expression data from microarray experiments, the authors discover 32 putative regulatory elements and 55 putative regulatory modules, demonstrating the effectiveness of their techniques.

Finally, in "Clustering gene expression data using an effective dissimilarity measure," Rosy Das *et al.* propose a novel dissimilarity measure and two clustering methods based on the dissimilarity measure. As demonstrated by the authors, their methods are effective in analysing gene expression data and are comparable to some well-known clustering algorithms in terms of homogeneity, silhouette, and the z-score cluster validity measure.

The goal of this special issue is to introduce the reader to the interdisciplinary field combining data mining, pattern analysis, and computational bioscience. We hope the conference report and articles in this special issue are interesting and suggest some directions for future research in this rapidly growing field.

Sanghamitra Bandyopadhyay
Ujjwal Maulik
Jason T. L. Wang