Guest Editorial
Foreword to the Special Issue on Intelligent Computation for Bioinformatics

THE LAST few decades have witnessed significant advancements in intelligent computation techniques. Driven by the need to solve complex real-world problems, powerful and sophisticated intelligent data analysis technologies have been exploited or emerged, such as neural networks, support vector machines, evolutionary algorithms, clustering methods, fuzzy logic, particle swarm optimization, data mining, etc. In recent years, the volume of biological data has been increasing exponentially, thus, allowing significant learning and experimentation to be carried out using a multidisciplinary approach, which gives rise to many challenging problems. Bioinformatics has become an ideal research area where computer scientists can apply and further develop new intelligent computation methods, in both experimental and theoretical cases.

The overall aim of this special issue is to bring together the latest applications and reviews of innovative intelligent computation tools for processing, management, analysis, interpretation, and integration of biological information. There were more than 20 papers submitted to this special issue, which covered both the practical and theoretical aspects of bioinformatics. After a rigorous peer-review process, six papers have been selected that emphasize what kind of intelligent computation tools are applied, and how these tools are exploited to solve the addressed biological data analysis problem.

Medical databases have accumulated large quantities of information about patients and their clinical conditions. Relationships and patterns hidden in this data can provide new medical knowledge as has been proved in a number of medical data mining applications. However, the data are rarely provided in a format suitable for immediate application of conventional attribute-valued learning (AVL). In the paper "Sequential Data Mining: A Comparative Case Study in Development of Atherosclerosis Risk Factors" by Kléma *et al.*, the problem of mining temporal and sequential medical data, which usually asks for complex preprocessing, is investigated. The sequences of the events are determined where each event is described by a numeric or symbolic value and a time stamp. The event types are also shown to be distinguished. The dataset can either be a single sequence, or it can be composed of a number of sequences. Strong sequential patterns are identified, i.e., such as event chains (subsequences) that appear frequently in the dataset, and their interaction is optionally studied with the target event.

Microarrays are at the center of a revolution in biotechnology, allowing researchers to simultaneously monitor the expression of tens of thousands of genes. Independent of the platform and the analysis methods used, the result of a microarray experiment is, in most cases, a list of genes found to be differentially expressed in different types of tissues. A common challenge faced by the researchers is to translate such gene lists into a better understanding of the underlying biological phenomena. In the paper "Learning Relational Descriptions of Differentially Expressed Gene Groups" by Trajkovski *et al.*, a new method is presented to identify the groups of differentially expressed genes that have functional similarity in the background knowledge, formally represented with gene annotation terms from the gene ontology. The input to the algorithm is a multidimensional numerical data set, representing the expression of the genes under different conditions (that define the classes of examples), and an ontology used for producing background knowledge about these genes. The output is a set of gene groups whose expression is significantly different for one class compared to the other classes. The features describe the differentially expressed genes in terms of their functionality and interactions with other genes.

The genome of several species has been available on the Internet for a couple of years now, which has resulted in a revolution in the bioinformatics field. Because of the availability in an electronic format, computers can be used to do experiments that would normally take months or even years. Data-mining research explores the raw data, identifying genes, designing primers, and searching for differences between individuals or differences between species. In the paper "Selection of DNA Markers" by Hoogeboom *et al.*, the problem of finding short (dis)similar substrings is studied for a given a genome, i.e., a long string over a fixed finite alphabet. An algorithm is first presented to detect the substrings that have edit distance to a fixed substring at the most equal to a given e. A second algorithm is then described that finds the set of all substrings having an edit distance larger than e to all others. Several applications are given, where attention is paid to practical biological issues such as hairpins and guanine–cytosine (GC) percentage, and an experiment shows the potential of the methods.

Microarray image processing consists of several steps, of which the first critical step is referred to as addressing or gridding. This is the process of identifying the areas within an image that contain a single spot, and identifying which subgrid and then which row and column within that subgrid the spot belongs to. In the paper "Blind Microarray Gridding: A New Framework" by Morris, the author demonstrates that the parameter-free gridding of a microarray image is possible, and implements a subgrid detection algorithm that has proven to work with real subgrids of various sizes. The approach toward gridding differs significantly from most of the other gridding frameworks in the literature as it does not make use of 1-D projections at any stage. The concept of regular-spaced grid fitness is taken into the gridding

Digital Object Identifier 10.1109/TSMCC.2007.906053

approach. Rather than simply trying to identify the number of rows and columns within the grid, the approach includes a measure of fitness for possible grids. By attempting to minimize this fitness value, a proven measure of consistency is brought to gridding across multiple images. Since the definition of regular grid fitness is accepted as credible, a method is introduced for comparing the accuracy between existing gridding techniques.

The rapid advances in high-throughput technologies such as DNA microarray have resulted in a great demand for visualizing multidimensional expression data in an effective way so that interesting patterns, features, and relationships can be extracted from the large data set. Visualization of high-dimensional data involves a combination of structural modeling and graphical representations. In the review paper "Information Visualization for DNA Microarray Data Analysis: A Critical Review" by Zhang *et al.*, it is explained how graphical representation can be applied in general to this problem domain, followed by exploring the role of visualization in gene expression data analysis. Having set the problem scene, the paper then examines various multivariate data visualization techniques that have been applied to microarray data analysis. These techniques are critically reviewed so that the strengths and weaknesses of each technique can be tabulated. Finally, several key problem areas as well as possible solutions to them are discussed as being a source for future work.

Identifying genes from DNA sequences is an important problem in bioinformatics. A computational approach to gene finding has recently attracted a lot more attention in the molecular biology and genomics community. In the review paper "Gene Identification: Classical and Computational Intelligence Approaches" by Bandyopadhyay *et al.*, the problem of gene identification, along with the issues involved in it, are first described. The classical approaches based on the hidden Markov model, Bayesian networks, and dynamic programming are then discussed. Finally, a review of some of the computational intelligence techniques for this problem is provided. In particular, the recent developments in various gene-finding methods, especially using computational intelligence techniques like neural networks and genetic algorithms, are summarized in this review, and a brief history as well as a description of these methods are provided. Web addresses for most of the gene-finding software and a fairly extensive bibliography are also included. Some of the limitations of the current methods are mentioned.

This special issue is a timely reflection of the research progress in the area of intelligent computation for bioinformatics. Finally, we like to acknowledge all the authors for their efforts in submitting high-quality papers. We are also very grateful to the reviewers for their thorough and on-time reviews of the papers. Last, but not the least, our deepest gratitude goes to the Editor-in-Chief and the Editorial Assistant of this TRANSACTIONS for their patience and great help.

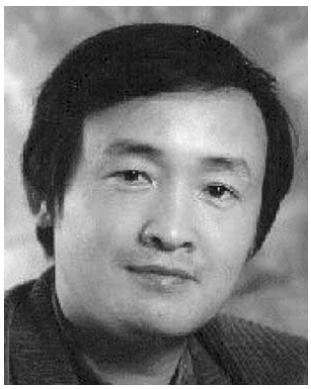
ZIDONG WANG, *Guest Editor*
School of Information Sciences and Technology
Donghua University
Shanghai 200051, China

Department of Information Systems and Computing
Brunel University
London UB8 3PH, U.K.

XIACHUI LIU, *Guest Editor*
Department of Information Systems and Computing
Brunel University
London UB8 3PH, U.K.

Zidong Wang (M'03–SM'04) is a Professor with Brunel University, London, U.K., and is also an Adjunct Professor with Donghua University, Shanghai, China. His research interests include dynamical systems, signal processing, bioinformatics, and control theory and applications. He has authored more than 100 papers published in refereed international journals.

Prof. Wang is currently an Associate Editor or Editorial Board Member for nine international journals including four IEEE TRANSACTIONS. He was awarded research fellowships from Germany, Japan, and Hong Kong.



Xiaohui Liu is a Professor of computing with Brunel University, London, U.K., where he directs the Centre for Intelligent Data Analysis. His current research interests include effective analysis of data, particularly in biomedical areas. He has authored over 180 refereed publications in data mining, bioinformatics, intelligent systems, and time series.

Prof. Liu is a Charted Engineer, Life Member of the Association for the Advancement of Artificial Intelligence, Fellow of the Royal Statistical Society, and Fellow of the British Computer Society.