Guest-editorial

Evolutionary computation in bioinformatics

Bioinformatics research, develop and apply computational tools and approaches for analyzing the use of biological, medical, behavioral and health data; and thus lies at the interface between biology, computer science and information technology. Bioinformatics, proteomics, metabolomics, genomics, molecular biology, molecular evolution, computational biology, and related fields are at the intersection between two axes: biological data analysis and information technology. Research in these fields comprises property abstraction from the biological system, design and development of data analysis algorithms, as well as of databases and web-based data access tools. The ultimate goal of the field is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology could be discerned. Bioinformatics is becoming popular in several AI-related areas, including evolutionary computation (EC). The main motivation for applying EC to bioinformatics related tasks is that they are robust and adaptive search methods, which perform a global search in the space of candidate solutions. The evolutionary computing community has been publishing bioinformatics-related articles in a relatively scattered manner in conference proceedings/journals dedicated to bioinformatics or evolutionary computing.

The objective of this issue was to assemble a set of high-quality original contributions that reflect some advances in the state-of-the-art in the area of evolutionary computation in bioinformatics; thereby presenting a consolidated view to the interested researchers in the aforesaid fields, in general, and readers of the Journal of Intelligent and Fuzzy System, in particular. It has four articles, each showing some advantages of EC and other related meta-heuristics for biological data analysis. The first one deals with combining support vector machine (SVM) based classifiers for biomedical data, while the second one deals with clustering of microarray gene profile data. The third article is dedicated to an

application of EC in protein-ligand docking; and the fourth article finds out an application of Ant Colony Optimization (ACO) in phylogenetic tree construction. Experts from different active groups from the U.S.A, Korea, Brazil and India author these articles; and each of them is reviewed by two to three referees.

Classification of biomedical data faces a special challenge because of the characteristics of the data: too few examples with too many features. Thus improving the classification performance or the generalization ability of a classifier in the biomedical domain is one of the active research areas. One solution to this problem is to build a fusion model to combine multiple classifiers together and result in a combined classifier that can achieve a better performance than any of its composing individual classifiers. In the first article of this issue, Chen et al. proposed a classifier fusion model to combine multiple support vector machines by employing fuzzy logic and genetic algorithms. The fuzzy logic system (FLS) is constructed based on SVM accuracies and distances of data examples to SVM hyper-planes. A genetic algorithm (GA) is used to tune the fuzzy membership functions in the FLS and determine the optimal fuzzy fusion model. They have reported promising results of the proposed model on two biomedical data sets namely colon tumor data and ovarian cancer data.

Clustering analysis of the gene expression profiles has been used for identifying the functions of unknown genes. Fuzzy clustering methods assign one sample to multiple clusters as their degrees of membership. It is shown to be more appropriate for analyzing gene expression profiles because genes usually belong to multiple functional families. However, general clustering methods have problems that they are sensitive to initialization and can be trapped into local optima. In this article Park and Cho proposed an evolutionary fuzzy clustering method with Bayesian validation that uses a genetic algorithm for fuzzy clustering process of gene

expression profiles and Bayesian validation method for the fitness evaluation process. They have conducted in-depth experiments to verify the usefulness of the proposed method with well-known gene expression profiles of SRBCT and Saccharomyces.

In the third paper Ghosh and Sen showed the utility of hybrid GAs to solve the flexible ligand docking problem i.e., predicting the binding conformation of a flexible ligand molecule into a rigid protein – a computationally hard problem in proteomics. The proposed hybrid GA scheme uses the concept of Lamarckian genetics to perform a local search about an individual, followed by replacing it with a better solution found in its neighborhood. Two local search schemes have been investigated and their performance relative to the standard GA have been compared. Preliminary results obtained on a set of three protein-ligand complexes have shown promising performance.

Another important problem in bioinformatics is the reconstruction of phylogenetic trees. A phylogenetic tree aims at unveiling the evolutionary relationship between several species. In this way, it is possible to know which species are more closely related to one another and which are more distantly related. Established methods for phylogeny work are good for small or moderate number of species, but they become unfeasible for large-scale data sets. In this article, Lopes and Perretto, proposed a methodology using ACO paradigm for the problem. A phylogenetic tree is viewed as a fully connected graph using a matrix of distances between species. They found out the shortest path in this graph using ACO. Two data sets were used to test the proposed model. The first one was used to investigate the sensitivity of the control parameters and to define their default values. The second data set was used to analyze the scalability of the system for a large number of sequences. The authors claimed that the proposed method is as good as or even better than the other conventional methods and very efficient for large-scale phylogeny.

Finally, we take this opportunity to thank Prof. R. Langari, Chief Editor, *Journal of Intelligent and Fuzzy Systems*, for giving us an opportunity to act as the Guest Editors for this special issue. We believe, the issue is very timely. We are thankful to all the contributors and reviewers for their co-operation in making this issue a reality.

The Guest-editors
Ashish Ghosh
Udo Seiffert
Lakhmi C. Jain
November 2006