
Preface

Biological systems are inherently stochastic and uncertain. Thus, research in bioinformatics, where computer technologies are applied to the management of biological data, and in computational biology, where computational models are built for modeling and analysis of ecological, molecular, cellular and neural networks, has to deal with a large amount of uncertainties. For example, a small number of molecules, different internal states of a population of cells, changes in environments and genetic mutations can all contribute to variations in gene expression.

Fuzzy logic has shown to be a powerful tool in capturing different uncertainties in engineering systems. In recent years, fuzzy logic based modeling and analysis approaches are becoming popular in analyzing biological data and modeling biological systems. Numerous results have been reported to demonstrate the effectiveness in applying fuzzy logic to solving a wide range of problems in bioinformatics, biomedical engineering, and computational biology.

This book contains 16 chapters that represent a body of selected research work on applying fuzzy systems to the modeling and analysis of biological systems, in particular, to bioinformatics, biomedical engineering and computational biology.

In Chapter 1, a method for generating type-1 and type-2 fuzzy rules using artificial immune systems (AIS) is presented. A brief introduction to both AIS and type-2 fuzzy systems is provided. The potential application of AIS-based fuzzy systems to data mining in bioinformatics and biomedicine is discussed.

Chapter 2 describes a framework for performing assembly of genome sequences of both single and multiple organisms using fuzzy logic. It is shown that fuzzy logic improves the performance of genome sequence assembly by allowing for tolerance of inexactness or errors in fragment matching and enhances the classification of fragments belonging to different organisms with a divide-and-conquer strategy.

Chapter 3 develops a model-based approach to the analysis of proteobacterial genomes for promoter features that is able to account for the variability in sequence, location and topology intrinsic to differential gene expression. Authors decompose a feature into a family of models or building blocks, which maximizes

the sensitivity of detecting those instances that weakly resemble a consensus (e.g., binding site sequences) without decreasing the specificity. These features are treated using fuzzy assignments, which allow them to encode how well a particular sequence matches each of the multiple models for a given promoter feature.

A data-adaptive fuzzy filtering framework for processing of cDNA microarray images is presented in Chapter 4. The framework is designed to remove noise in cDNA microarray images that does not require for fuzzy rules, nor does it assume that the original cDNA signal is available. This is achieved by utilizing the inference engine in the form of transformed distance metrics between the cDNA vector-valued samples within the supporting window. In this way, the training or learning of the weighting coefficients is only based on local image features without the use of linguistic fuzzy rules or local statistics estimation.

In Chapter 5, the fuzzy c-means clustering algorithm is suggested for analyzing microarray gene expression data followed by a discussion of the main concerns in clustering gene expression data. Tuning of the parameters are discussed in the context of 2-way and 3-way microarray data. A transformation that allows for more contrast in distances between all pairs of samples in a dataset is proposed, which increases the likelihood of detection of a group structure in a high dimensional dataset.

Chapter 6 introduces a flexible framework for feature selection and classification of microarray data. Dimensionality reduction is achieved by the application of a supervised fuzzy pattern algorithm that is able to reduce and discretize existing gene expression profiles. Then, a self-organizing neural network, termed growing cell structures (GCS) network, is employed for clustering biological data.

In Chapter 7, authors employ a fuzzy rule-based classification system to analyze gene expression data. The applied classifier consists of a set of fuzzy if-then rules that allow for efficient and accurate classification of input patterns. Furthermore, a hybrid fuzzy approach to classifying gene expression data, where a genetic algorithm is used to select a subset of the fuzzy rules, is also presented. It is shown that the performance of the compact fuzzy classifier is comparable to that of the full classifier.

Reconstruction of gene regulatory networks using fuzzy logic based models is reviewed in both Chapters 8 and 9. Chapter 8 emphasizes the functionalities of regulatory motifs, and the reconstruction of such motifs using fuzzy systems. Two selected methods are discussed in detail, examples are given where the two methods are applied to both real microarray data concerning the yeast cell cycle and simulated data concerning the Raf signaling pathway. In contrast, Chapter 9 surveys the application of fuzzy logic both to clustering of gene expression data and the reconstruction of gene regulatory networks. Examples are also supplied in the discussion of the methods.

Chapters 10, 11 and 12 provides examples of employing fuzzy logic to model gene expression data and biological networks in greater detail. Chapter 8 describes the use of genetic programming to evolve a fuzzy rule base to model gene expression. It is shown that fuzzy rule based models allow for the insertion of

prior knowledge, which makes it possible to find sets of rules that include the relationships between genes that are already known. In addition, it is demonstrated that evolving a fuzzy rule base using genetic programming is able to extract explanatory rules from microarray data obtained in the real experiments.

Chapter 11 describes a class of widely used neuro-fuzzy systems, known as adaptive neuro-fuzzy inference systems (ANFIS), and its application to modeling gene regulatory networks. Furthermore, a hierarchical, multi-layer ANFIS model (termed GeneCFE-ANFIS) is introduced. It is shown that GeneCFE-ANFIS is able to improve the performance of prediction in terms of true positive rate with a little amount of a priori knowledge about gene interactions.

Although fuzzy logic presents an appealing approach to modeling gene expression data, it also faces the serious challenge of combinatorial rule explosion in modeling complex biological networks. Chapter 12 suggests a number of approaches to addressing the scalability issue, including adopting the union rule configuration or optimizing the fuzzy rule structure using genetic algorithms.

Chapters 13, 14 and 15 provide various applications of fuzzy logic to dealing with biomedical problems. In Chapter 13, authors provide an overview of several fuzzy c-means based clustering approaches to medical imaging. The conventional hard c-means and the fuzzy c-means, together with three computationally more efficient variants of fuzzy c-means are evaluated. In Chapter 14, the application of self-organizing fuzzy logic controller (SOFLC) to the control of a multivariable model of anesthesia is explored. A methodology is proposed to design SOFLC for complex multi-input/multi-output (MIMO) systems. Different design strategies of MIMO are outlined and the application of such SOFLC systems to muscle relaxation and depth of anesthesia control is studied. Chapter 15 presents an interval type-2 fuzzy classifier and its application to ECG arrhythmic classification problem. It is shown that the uncertainties associated with the membership functions can be encapsulated by the footprint of uncertainty (FOU) and that it can be fully characterized by the upper membership function (UMF) and lower membership function (LMF). The proposed type-2 fuzzy classifier is applied to the ECG arrhythmic classification problem and the performance of the classifier is tested on MIT-BIH Arrhythmia database. Results show that the proposed strategies to design the FOU are essential to achieve a high performance fuzzy rule-based classifier in the presence of uncertainties.

The role of fuzzy logic in the modeling of gene regulatory networks is further studied in Chapter 16. Different to its applications in modeling and analysis of gene expression data, this chapter investigates *in silico* the influence of control logic on the easiness of evolving typical regulatory dynamics in computational models of genetic regulatory networks. The gene regulatory network motif considered in this work consists of three genes with both positive and negative feedback loops. Two fuzzy logic formulations are studied in this work, one is known as the Zadeh operator, and other is the probabilistic operator. Empirical results show that with the probabilistic ‘AND’ operator and the probabilistic ‘OR’ operator, the system is able to evolve sustained oscillation with a low probability. However, sustained oscillation is not evolvable when the Zadeh operator is

employed. In addition, it is also shown that regulatory motifs with the probabilistic operators possess much richer dynamics than that with the Zadeh operators.

The research work described in this book presents a selected yet comprehensive picture of how fuzzy logic can elegantly address problems in bioinformatics, biomedical engineering and computational biology, particularly in dealing with uncertainties in biological systems. We hope that the methodologies and application examples discussed in the book are instructive and inspiring to both practitioners and researchers. Thus, we hope that the publication of this book will further promote the related research areas.

We would like to thank Dr. Janusz Kacprzyk for including this book in the Springer book series “Studies in Fuzziness and Soft Computing”. We are also grateful to the authors for their nice contributions and cooperation during the preparation of the book. Finally, we would like to thank Heather King and Thomas Ditzinger of Springer for their kind support and patience.

October 2008

Yaochu Jin
Lipo Wang

Fuzzy Systems in Bioinformatics and Computational
Biology

Jin, Y.; Wang, L. (Eds.)

2009, XVI, 332 p. 118 illus., 11 illus. in color., Hardcover

ISBN: 978-3-540-89967-9