

Editorial

Genomic Signal Processing

Ulisses Braga-Neto,¹ Rui Kuang,² Harri Lähdesmäki,³ Haris Vikalo,⁴ and Byung-Jun Yoon¹

¹Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA

²Department of Computer Science and Engineering, University of Minnesota Twin Cities, Minneapolis, MN 55455, USA

³Department of Information and Computer Science, Aalto University, Espoo 00076, Finland

⁴Department of Electrical and Computer Engineering, University of Texas, Austin, TX 78712, USA

Correspondence should be addressed to Ulisses Braga-Neto, ulisses@ece.tamu.edu

Received 19 December 2010; Accepted 19 December 2010

Copyright © 2010 Ulisses Braga-Neto et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Genomic signal processing (GSP) is the engineering discipline that aims to integrate the theory and methods of signal processing with the applications arising from high-throughput technologies in biomedical research, such as gene-expression microarrays or proteinabundance mass spectrometry. GSP comprises the analysis, processing, and use of genomic and proteomic signals for gaining knowledge into the complex structural and functional relationships among genes and proteins in living tissue, as well as the translation of that knowledge into systems-based medical applications. GSP has had a significant impact in biomedical research over the last decade and promises to revolutionize medical practice.

This special issue consists of 8 papers covering a broad range of topics in bioinformatics, statistical signal processing, stochastic modeling, pattern recognition, and systems identification, with applications in target estimation in microarray signals, gene structure identification, and error estimation for discrete gene prediction, inference of gene regulatory networks, classification of phenotypic data, and transcription factor binding site prediction. Next we will briefly introduce the papers comprising this special issue.

H. Vikalo and M. Gokdemir model binding in real-time DNA microarrays as a stochastic differential equation, and they proposed a solution method based on a Markov-Chain Monte-Carlo (MCMC) technique. The authors found that the proposed technique significantly outperformed previously proposed methods in simulation experiments. They also tested the proposed approach on experimental data.

S. Winter-Hilt and C. Baribault introduce a generalized-clique hidden Markov model (HMM) in order to improve the modeling of the critical signal information at the transitions between exon regions and noncoding regions of DNA. The authors applied the proposed model to gene finding in *C. elegans*, and showed that the individual-state and full-exon predictions are greatly enhanced over the standard HMM when using the generalized-clique HMM.

The paper by S. Winter-Hilt et al. describes a new method to introduce duration into an HMM, by using side information, that is, put in the form of a martingale series. Their method allows the HMM to use fully the side-information available during its dynamic table optimization, such as in Viterbi path calculations.

T. Chen and U. Braga-Neto present analytical formulations of the bias, variance, and RMS of coefficient of determination (CoD) estimators for discrete prediction. The authors present numerical experiments that compare the performance of several CoD estimators and conclude that the empirical (resubstitution) CoD can be the best choice, provided that one has evidence of moderate to tight regulation between the genes, and the number of predictors is not too large.

J. Meng et al. propose a novel Bayesian sparse-correlated rectified factor model (BSCRFM) that models the unknown transcription factor (TF) protein level activity, the correlated regulations between TFs, and the sparse nature of TF-regulated genes. The method utilizes prior information from existing databases and performs inference of gene regulatory networks through a Gibbs sampling algorithm that

the authors developed. The technique was validated using both simulated data and actual microarray data from breast cancer patients.

The paper by T. Vu and U. Braga-Neto investigates the performance of various error estimation methods for bagged ensemble classification rules based on small sample data. The authors proposed an explicit definition of the out-of-bag estimator, that is, intended to remove estimator bias. Using both synthetic data and data from published gene-expression and protein-abundance studies, the authors report that the out-of-bag estimator is similar in performance to the leave-one-out estimator, the performance of other estimators is consistent with what has been observed in the previous studies using single classifiers, and bolstered error estimators showed the best performance.

X. Shen and H. Vikalo propose a particle filter with a Markov-Chain Monte-Carlo move step for the estimation of reaction rate constants in gene regulatory networks modeled by chemical Langevin equations. The authors report that the proposed technique outperformed previously considered methods, while being computationally more efficient. In addition, the authors compute an approximation to the Cramer-Rao lower bound on the mean-square error of estimating reaction rates and demonstrate that when the number of unknown parameters is small, the proposed particle filter can be nearly optimal.

Finally, the paper by X. Dai et al. develops a new data fusion method to combine multiple genome-level data sources for transcription factor binding predictions. Using a carefully constructed test set of verified binding sites in the mouse genome, the authors report that their method can reduce false positive rates, and that DNA duplex stability and nucleosome occupation data can improve the accuracy of transcription factor target gene predictions. The authors conclude that that nonredundant data sources provide the most efficient data fusion.

Acknowledgments

The authors would like to thank the EURASIP JASP Editor-in-Chief, Dr. Phillip Regalia, for the opportunity to coordinate this special issue, and the anonymous reviewers for their diligent work, without which this special issue would not have been possible.

Ulisses Braga-Neto
Rui Kuang
Harri Lähdesmäki
Haris Vikalo
Byung-Jun Yoon