Guest Editorial: Introduction to the Special Issue on Machine Learning for Microarray Bioinformatics

MAN WAI MAK

The Hong Kong Polytechnic University, Kowloon, Hong Kong

AHMED TEWFIK

University of Minnesota, Minneapolis, MN, USA

LAI WAN CHAN

Chinese University of Hong Kong, Sha Tin, Hong Kong

CHUN CHUNG KEITH CHAN

The Hong Kong Polytechnic University, Kowloon, Hong Kong

One of the main challenges in computational biology is the revelation and interpretation of the rich genomic information underlying cancer biology and to facilitating molecular classification and prediction of cancers and responses to therapies. Genomic sequencing and gene expression technologies have been widely recognized as vital approaches to modern drug design and disease classification. With the recent advances in DNA microarray technologies, it has become possible to measure the expression level of thousands of genes simultaneously. However, the large number of genes together with the complexity of gene expression patterns and sequences make interpreting the million of biological measurements a challenging task. This special issue showcases various machine learning approaches to meeting this challenge. The areas in which this special issue is focusing on can be categorized as follows.

1. *Clustering of gene expression data*. Clustering of gene expression data has been a popular approach to identifying groups of genes or groups of samples that exhibit similar expression patterns. The results of clustering have a wide range of applications, including disease diagnoses, gene identification and discovery, etc. In "Teng and Chan, *Discovering Biclusters by Iteratively Sorting with Weighted Correlation Coefficient in Gene Expression Data*," a framework for biclustering gene expression profiles is proposed. The biclustering algorithm alternatively sorts the genes and conditions using dominant sets, resulting in multiple overlapping biclusters at the end of the iterative process. A new criterion for measuring the homogeneity of biclusters is also proposed. In "Lam et al., *A Regularized Clustering Algorithm Based on Calculus of Variations*," the descent direction of the l_{2m} fuzzy C-means algorithm is constrained via calculus of variations in order to avoid trapping in local minima during the clustering process.

2. *Gene selection*. A microarray gene expression matrix typically contains thousands of genes but only tens or hundreds of samples. Classification of the samples using all of the genes as features will likely be hampered by the curse of dimensionality problem. Fortunately, only a small number of genes are informative for classification. Therefore, it is imperative to derive a criterion for selecting the most informative set of genes. In the paper "*Effective Gene Selection*

Method using Bayesian Discriminant Based Criterion and Genetic Algorithms," Gan et al. propose a new gene selection criterion based on the concept of Bayesian discriminant.

3. *Spot Segmentation*. The reliability of clinical applications of microarrays depends on the accuracy of the gene expression level extracted from the spots on the microarrays, which in turn relies on the accuracy of spot segmentation. In "*A Comparison of Fuzzy Clustering Approaches for Quantification of Microarray Gene Expression*," Wang et al. propose a fuzzy-clustering spot-segmentation method that can handle array spots with complex shapes such as donuts and scratches.

4. *Dynamic of gene expression*. Understanding the dynamic processes of gene expression can help optimize biological process for the biotech industry. In "*Projecting Gene Expression Trajectories through Inducing Differential Equations from Microarray Time Series Experiments*," Kramer and Xu describe a method for inferring a continuous model from discrete gene measurements. The inferred model can help predict biological pathways, which is essential for many biological applications.

5. *Inferring Gene Networks*. Recently there has been much interest in inferring gene and protein interaction networks. In the paper "*Enhancing Automatic Construction of Gene Subnetworks by Integrating Multiple Sources of Information*," Suwannaroj and Niranjan present an approach to inferring cellular gene pathways based on the textual information extracted from literatures.

We are grateful to the authors and reviewers who contributed to the papers in this special issue.

Man-Wai Mak received his BEng(Hons) degree in Electronic Engineering from Newcastle Upon Tyne Polytechnic and a PhD degree in Electronic Engineering from the University of Northumbria at Newcastle. He joined the Department of Electronic Engineering at the Hong Kong Polytechnic University as a Lecturer in 1993 and as an Assistant Professor in 1995. He has authored more than 90 technical papers in speaker recognition, machine learning, and bioinformatics. Dr. Mak is also a co-author of the postgraduate textbook "Biometric Authentication: A Machine Learning Approach, Prentice Hall, 2005." Dr. Mak received the Faculty of Engineering Research Grant Achievement Award in 2003. Since 1995, Dr. Mak has been an executive committee member of the IEEE Hong Kong Section Computer Chapter. He was the Chairman of the IEEE Hong Kong Section Computer Chapter in 2003-2005 and is currently a member of the IEEE Machine Learning for Signal Processing Technical Committee. He served as a guest editor of a special issue of Journal of VLSI Signal Processing Systems, Pattern Analysis and Applications, and International Journal of Computational Intelligence Research. Dr. Mak's research interests include speaker recognition, machine learning, and bioinformatics.

Ahmed H. Tewfik was born in Cairo, Egypt on October 21, 1960. He received his B.Sc. degree from Cairo University, Cairo, Egypt, in 1982 and his M.Sc., E.E., and Sc.D. degrees from the Massachusetts Institute of Technology, Cambridge, MA, in 1984, 1985, and 1987, respectively. Dr. Tewfik has worked at Alphatech, Inc., Burlington, MA in 1987. He is the E. F. Johnson professor of Electronic Communications with the Department of Electrical Engineering at the University of Minnesota. He served as a consultant to MTS Systems, Inc., Eden Prairie, MN, Emerson-Rosemount, Inc., Eden Prairie, MN; CyberNova, Milipitas, CA; Macrovision, Santa Clara, CA, and Ipsos, New York. He worked with Texas Instruments and Computing Devices International. From August 1997 to August 2001, he was the President and CEO of Cognicity, Inc., an entertainment marketing software tools publisher that he co-founded, on partial leave of absence from the University of Minnesota. His current research interests are in signal processing for high performance local and personal area wireless networks, pervasive datanomic computing, multimedia and genomics. Professor Tewfik is a fellow of the IEEE. He was a distinguished lecturer of the IEEE Signal Processing Society in 1997-1999. He received the IEEE third Millennium award in 2000. He was invited to be a principal lecturer at the

1995 IEEE EMBS summer school. He was awarded the E. F. Johnson professorship of Electronic Communications in 1993, a Taylor faculty development award from the Taylor foundation in 1992, and an NSF research initiation award in 1990. He delivered plenary lectures at several IEEE and non-IEEE meetings, including the 1994 IEEE International Conference on Acousting Speech and Signal Processing. (ICASSP94), the 1999 IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing, the 1999 IEEE Turkish Signal Processing Conference (SIU99), the 1st IEEE International Symposium on Signal Processing and Information Theory (2001), SSGRR2002w International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet, the 2003 European Union COST meeting, and the 10th IEEE International Conference on Electronics, Circuits and Systems. He gave invited tutorials on ultra-wideband communications at the 2003 Fall IEEE Vehicular Technology Conference, watermarking at the 1998 IEEE International Conference on Image Processing and Wavelets at the 1994 IEEE workshop on Time-Frequency and Time-Scale Analysis. He was selected to be the first Editor-in-Chief of the IEEE Signal Processing Letters from 1993 to 1999. He is a past associate editor of the IEEE Transactions on Signal Processing, was a guest editor of three specialissue of that journal on wavelets and their applications and watermarking and a guest editor of a special issue of the IEEE Transactions on Multimedia on multimedia databases.

financial engineering and bioinformatics to perform temporal and spatial data analysis. Applications include portfolio management, risk analysis, trading systems, gene expression data analysis and time series analysis. Other research works include the learning and modelling of the feedforward and the recurrent networks, the modular learning methods, and applications of neural networks in image recognition, Cantonese speech recognition and multimedia database. Professor Chan has published more than 100 scientific papers in the above research areas.

Prof. Keith Chan obtained his B.Math. (Hons.) degree in Computer Science and Statistics, and M.A.Sc. and Ph.D. degrees in Systems Design Engineering from the University of Waterloo, Waterloo, Ontario, Canada. He has a number of years of academic and industrial experience in software development and management. He joined the IBM Canada Laboratory, Toronto Canada, in 1989, where he was involved in the development of multimedia and software engineering tools. In 1993, he joined the Department of Electrical and Computer Engineering at Ryerson University, Toronto, Ontario, Canada as an Associate Professor. He returned to HK in 1994 to join the Hong Kong Polytechnic University where he is currently a Professor and Head of the Department of Computing. He is currently also a Guest Professor of the Graduate University of the Chinese Academy of Sciences. Prof. Chan has co-authored two books on Agile Software Development and Software Process Improvement respectively. He has recently completed another book on Software Development Rhythm to be published by John Wiley. Prof. Chan has over 100 technical publications in the areas of Software Engineering, Data Mining and Bioinformatics. He has served on the program committees of various conferences and has also provided consultancy services to government agencies and large and small to medium sized companies in Hong Kong, China, Singapore, Malaysia, and Italy.

Prof. Lai-wan Chan received her BA, MA, and PhD degrees in Engineering from the University of Cambridge. She is currently a Professor in the Computer Science and Engineering Department, and the Associate Dean (Education) in the Faculty of Engineering at the Chinese University of HongKong. Her research interests is in data mining, financial engineering, bioinformatics and artificial neural networks. She has designed and applied various neural networks and ICA techniques in