2009 International Joint Conferences on Bioinformatics, Systems Biology and Intelligent Computing

IJCBS 2009

SHANGHAI, CHINA
3 - 5 AUGUST 2009

Editors:
Joe Zhang, Guozheng Li, Jack Y. Yang

Associate Editors:
Abdullah Arslan, Sushing Chen, Youping Deng, Yufei Huang, Armin R. Mikler, Yufeng Wang, Dan Xi, Zhongming Zhao

Proceedings

2009 International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing

IJCBS 2009

3-5 August 2009
Shanghai, China

CPS
Conference Publishing Services

Los Alamitos, California
Washington • Tokyo

IEEE computer society

IEEE Computer Society Order Number P3739
ISBN-13: 978-0-7695-3739-9
BMS Part # CFP0935H-PRT
Library of Congress Number 2009904091

Additional copies may be ordered from:

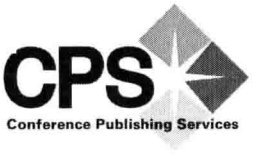
IEEE Computer Society
Customer Service Center
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1314
Tel: + 1 800 272 6657
Fax: + 1 714 821 4641
http://computer.org/cspress
csbooks@computer.org

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
Tel: + 1 732 981 0060
Fax: + 1 732 981 9667
http://shop.ieee.org/store/
customer-service@ieee.org

IEEE Computer Society
Asia/Pacific Office
Watanabe Bldg., 1-4-2
Minami-Aoyama
Minato-ku, Tokyo 107-0062
JAPAN
Tel: + 81 3 3408 3118
Fax: + 81 3 3408 3553
tokyo.ofc@computer.org

Individual paper REPRINTS may be ordered at: <reprints@computer.org>

Editorial production by Juan E. Guerrero
Cover art production by Mark Bartosik
Printed in the United States of America by Applied Digital Imaging



IEEE Computer Society
***Conference Publishing Services* (CPS)**
http://www.computer.org/cps

Proceedings

2009 International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing

IJCBS 2009

Message from the General Chair

It is an exciting time for Bioinformatics, Genomics and Systems Biology, a time of continuous celebration of seemingly endless series of revolutionary inventions, and powerful next generation sequencing technologies. It is an exciting time to develop new powerful computational methods for deciphering the flood of genomic data and lay down foundations of Genome Biology.

It is an exciting time for Bioinformatics in China, where the growing number of researchers in the field has reached a critical mass, the level at which it is right to say that China became a major player in Bioinformatics along with Europe, Japan and the US.

I am delighted to see a number of high level papers submitted to the Shanghai International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing, the papers describing innovative approaches to long standing problems in genomic sequence analysis, prediction of structure and biological function as well as applications to molecular, personalized and alternative medicine.

Mark Borodovsky
General Chair

Message from the Program Chairs

Welcome to The 2009 International Joint Conferences on Bioinformatics, Systems Biology and Intelligent Computing (IJCBS'09) being held in Shanghai, China on August 3-5, 2009. Bioinformatics, Systems Biology and Intelligent Computing are synergistic disciplines that hold great promise for the advancement of research and development in designing intelligent systems to solve systems biology and translational bioinformatics problems. The IJCBS'09 is aimed at providing a common platform to bridge these very important interdisciplinary areas into an interactive forum, and bringing together top researchers, practitioners and students from around the world in order to promote scientific understanding and findings in computing intelligence and bioscience. It is sponsored by IEEE Computer Society, International Society of Intelligent Biological Medicine (ISIBM), USA National Science Foundation (NSF), National Natural Science Foundation of China (NSFC), Tongji University, and University of Southern Mississippi, USA.

The program committee consists of more than 100 committee members around the world who served as workshop chairs, session chairs, and peer reviewers. The IJCBS'09 received more than 240 submissions through the online submission systems or emails. The international nature of the IJCBS'09 is reflected in the geographical diversity of the submission pool. All submitted papers have been peer-reviewed by the program committee members or invited external reviewers. A total of 52 papers have been selected for full papers and oral presentation with an acceptance ratio of 22%, and registered authors are from 17 countries including Australia, Belgium, Canada, China, Denmark, Germany, India, Iran, Japan, Jordan, Portugal, Singapore, South Korea, Spain, Turkey, UK and USA. A number of other papers have been accepted as short papers or posters. The conference features eight distinguished keynote speakers, a panel, best paper awards and travel fellowships.

Many individuals have contributed to the success of this conference. Many thanks go to all the authors, invited speakers and conference organizers with special thanks to the general chair Dr. Mark Borodovsky, honorary general chairs Dr. Ruzena Bajcsy, Dr. Michael Waterman, Dr. Joydeep Ghosh and Dr. Changjun Jiang, and all steering committee members whose leadership ensures the success of this conference. Special thanks go to the organizing chair Dr. Weisheng Xu, and local organizing committee consisting of Drs. Jing Yao, Mingyu You, Youling Yu, and Yongqing Su for their assistance.

The help and support from the IEEE Computer Society is especially appreciated. Andrea Thibault-Sanchez from IEEE Conference Publishing Services (CPS) provided us much advice and answered our inquiries. The IEEE Computer Society editor, Juan Guerrero, did an outstanding job in preparing the proceedings.

We wish everybody an enjoyable and fruitful stay in Shanghai.

Joe Zhang
Program Chair

Guozheng Li
Program Co-chair

Organizing Committee

Honorary General Chairs
Ruzena Bajcsy, *University of California, Berkeley, USA*
Joydeep Ghosh, *University of Texas, Austin, USA*
Michael S Waterman, *University of Southern California, Los Angles, USA*

General Chair
Mark Borodovsky, *Georgia Institute of Technology, USA*

Program Committee Chair
Joe Zhang, *University of Southern Mississippi, USA*

Program Committee Co-chairs
Andreas Dress, *CAS-MPG Partner Institute for Computational Biology, China*
Guo-Zheng Li, *Tongji University, China*

Publication Chairs
Heng Huang, *University of Texas, Arlington, USA*
Sumanth Yenduri, *University of Southern Mississippi, USA*
Zhongming Zhao, *Virginia Commonwealth University, USA*

Workshop Chairs
Abdullah N. Arslan, *University of Vermont, USA*
Ping Gong, *2SpecPro Inc., USA*

Special Session Chairs
Jacqueline Signorini, *University Paris 8, France*
Tianrui Li, *Southwest Jiaotong University, China*

Tutorial Chair
Shuigeng Zhou, *Fudan University, China*

Publicity Chairs
Jiaoxiong Xia, *Shanghai Municipal Education Commission, China*
Yilong Yin, *Shandong University, China*
Mehdi Pirooznia, *Johns Hopkins University, USA*

Registration Chair
Youping Deng, *University of Southern Mississippi, USA*
Fei Qiao, *Tongji University, China*

Organizing Chair
Weisheng Xu, *Tongji University, China*

Organizing Vice-Chairs
Mingyu You, *Shanghai University, China*
Preetam Ghosh, *University of Southern Mississippi, USA*

Award Chairs
Susan Bridges, *Mississippi State University, USA*
Hamid R. Arabnia, *University of Georgia, USA*
Yong-Sheng Ding, *Donghua University, China*

Local Chair
Jing Yao, *Tongji University, China*

Steering Committee

Hamid R. Arabnia

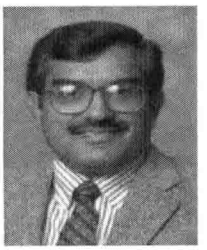
Founding chair of WORLDCOMP Congress, Editor-in-Chief of the Journal of Supercomputing (Springer) and recipient of William F. Rockwell, Jr. Medal for promotion of multi-disciplinary research (Rockwell edal is International Technology Institute's highest honor). He is on the editorial and advisory boards of 25 other journals and science magazines.

Ruzena Bajcsy

Member of United States National Academy of Science, Institute of Medicine, Member of United States National Academy of Engineering, 2003-2005 United States President's Information Technology Advisory Committee (PITAC), 2001 ACM Allen Newell Award, Member of National Institute of Standards and Technology, Member of National Research Council Army Research Technical Assessment Board, Member of Computer Research Association for Women, Member of the Review Board of Stanford University Computer Science, Fellow of ACM and Fellow of IEEE, Distinguished Professor and Director of CITRIS, University of California at Berkeley, USA.

Mark Borodovsky

President of International Society of Intelligent Biological Medicine, Founder of GeneMark and Pioneer of Bioinformatics Research, Founder of Georgia Tech Bioinformatics Ph.D. Programs, Distinguished Regents' Professor and Director Center for the Bioinformatics and Computational Genomics, Georgia Institute of Technology, USA.



Arif Ghafoor

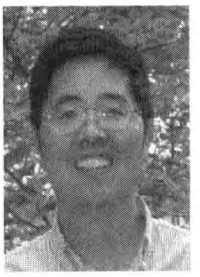
Professor of Electrical and Computer Engineering

Research:
Multimedia systems, databases,
distributed computing systems,
broadband multimedia networking

Areas of Interest:
Communications, Networking,
Signal & Image Processing
Computer Engineering

Joydeep Ghosh

Distinguished Schlumberger Centennial Chair Professor of Engineering, Chair of IEEE DMTC and Fellow of IEEE, 1992 IEEE Darlington Award, Director of IDEAL (Intelligent Data Exploration and Analysis Lab), University of Texas at Austin, U.S.A.

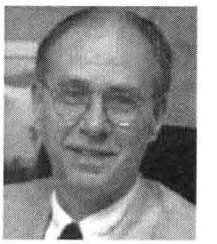


Tony Hu

IEEE Computer Society Bioinformatics and Biomedicine Steering Committee Chair and the IEEE Computational Intelligence Society Granular Computing Technical Committee Chair, the founding editor-in-chief of the International Journal of Data Mining and Bioinformatics, an associate editor/editorial board member of four international journals (KAIS, IJDWM, IJSOI and JCIB). He has received a few prestigious awards including the 2005 National Science Foundation (NSF) Career award.

Yi Pan

Editor-in-chief of International Journal of Bioinformatics Research and Applications, series editor of Wiley Book Series on Bioinformatics, Associate Editor of 6 IEEE Transactions, winner of many awards, keynote speaker for over 10 conferences, and Chair of Department of Computer Science at Georgia State University, USA

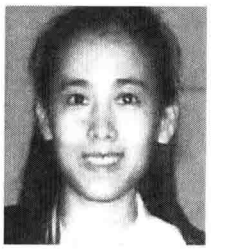


Michael S. Waterman

Member of United State National Academy of Science, Fellow of American Academy of Art and Sciences and Academician of French Académie des Sciences. Coeditor-in-Chief of Journal of Computational Biology, Editor-in-Chief of Lecture Notes in Bioinformatics, Lecture Notes in Computer Science (LNCS, Springer). Distinguished University Professor of University of Southern California, Los Angles, California, USA

Cathy H. Wu

Dr. Wu has conducted bioinformatics research since 1990 and developed several protein classification systems and databases. She has managed large software and database projects, led the bioinformatics effort of the Protein Information Resource (PIR) since 1999, and become the PIR Director in 2001. Her research interests include protein family classification and functional annotation, biological data integration, and literature mining.



Mary Qu Yang

Editor-in-Chief, International Journal of Computational Biology and Drug Design. Oak Ridge Institute for Science and Education, Oak Ridge National Lab. U.S. Department of Energy, USA National Human Genome Research Institute, National Institutes of Health (NIH) , U.S. Department of Health and Human Services, Bethesda, MD 20852 USA.

Aidong Zhang

Dr. Aidong Zhang is a professor in the Department of Computer Science and Engineering at the State University of New York at Buffalo and the director of the Buffalo Center for Biomedical Computing (BCBC). She is an author of more than 200 research publications and has served on many editorial boards of prestigious journals. Dr. Zhang is a recipient of the National Science Foundation CAREER Award and SUNY (State University of New York) Chancellor's Research Recognition Award. Dr. Zhang is an IEEE Fellow.

Program Committee

Sadaf Alam, *Oak Ridge National Laboratory, USA*
Laurence Rodrigues Do Amaral, *Federal University of Goias, Brazil*
Purushotham Bangalore, *University of Alabama at Birmingham, USA*
Pierre Beauseroy, *University of Technology in Troyes, UTT, France*
Mahua Bhattacharya, *Indian Institute of Information Technology & Management, India*
Hong Cai, *University of Texas at San Antonio, USA*
Rui Camacho, *Porto University, Portugal*
Zhiwei Cao, *Tongji University, China*
Cornelia Caragea, *Iowa State University, USA*
Doina Caragea, *Kansas State University, USA*
Rui Chang, *University of California, USA*
Bernard Chen, *University of Central Arkansas, USA*
Jake Y. Chen, *Purdue University, USA*
Yidong Chen, University of Texas Health Science Center at San Antonio, USA
Jianlin Cheng, *University of Missouri, USA*
Yanfeng Cheng, *Georgia Institute of Technology, USA*
Qiang Cheng, *Southern Illinois University Carbondale, USA*
Sung-Bae Cho, *Yonsei University, Korea*
Albert C. S. Chung, *Hong Kong University of Science and Technology, Hong Kong*
Mark Clement, *Brigham Young University, USA*
Kevin Daimi, *University of Detroit Mercy, USA*
Phuongan Dam, *University of Georgia, USA*
Yisheng Ding, *Donghua University, China*
Qiwen Dong, *Fudan University, China*
Pan Du, *Northwestern University, USA*
Ye Duan, *University of Missouri - Columbia, USA*
Werner Dubitzky, *University of Ulster at Coleraine, United Kingdom*
Minrui FEI, *Shanghai University, China*
Elisa Ficarra, *Politecnico di Torino, Italy*
Christopher M. Frenz, *New York City College of Technology (CUNY)*, USA
Jean X. Gao, *University of Texas at Arlington, USA*
Yang Gao, *Nanjing University, China*
Xin Geng, *South East University, China*
Preetam Ghosh, *University of Southern Mississippi, USA*
Kreshna Gopal, *Texas A&M University, USA*
Jianying Gu, *City University of New York, USA*
Jun-tao Guo, *University of North Carolina at Charlotte, USA*
Maozu Guo, *Harbin Institute of Technology, China*
Yufei Huang, *University of Texas at San Antonio, USA*
Wen-Lian Hsu, *Institute of Information Science, Academia Sinica, Taiwan*
Hongwei Huo, *Xidian University, China*
Jaeseung Jeong, *Korea Advanced Institute of Science and Technology, Korea*
Liangxiao Jiang, *China University of Geosciences, China*
Rui Jiang, *Tsinghua University, China*
Zhenran Jiang, East *China Normal University, China*
Jaewoo Kang, *Korea University, Korea*
Abdellali Kelil, *University of Sherbrooke, Canada*
Dongsup Kim, *Korea Advanced Institute of Science and Technology, Korea*
Hyunsoo Kim, *Harvard-Partners Center for Genetics and Genomics, USA*
Sun Kim, *Indiana University Bloomington, USA*
Jiejun Kong, *University of California, USA*
Wei Kong, *Shanghai Maritime University, China*
Rui Kuang , *Department of Computer Science and Engineering, University of Minnesota, USA*

Zoé Lacroix, *Arizona State University, USA*
Feipei Lai, *National Taiwan University, Taiwan*
Chang-Shing Lee, *National University of Tainan, Taiwan*
Haim Levkowitz, *University of Massachusetts Lowell, USA*
Chang-Tsun Li, *University of Warwick, United Kingdom*
Jia Li, *Shanghai University, China*
Kang Li, *Queen's University Belfast, United Kingdom*
Liao Li, *University of Delaware, USA*
Ming Li, *Nanjing University, China*
Shao Li, *Tsinghua University, China*
Shaozi Li, *Shamen University, China*
Shen Li, *Indiana University School of Medicine, USA*
Tianrui Li, *Southwest Jiaotong University, China*
Xiaoli Li, *Nanyang Technological University, Singapore*
Xiaolin Li, *Nanjing University, China*
Ying-Xin Li, *Nanjing University, China*
Tim. G. Lilburn, *American Type Culture Collection, USA*
Chun-Yuan Lin, *Chang Gung University*
Simon Lin, *Northwestern University Biomedical Informatics Center, USA*
Hongfang Liu, *Georgetown University, USA*
Jun Liu, *NUAA, China*
Qingzhong Liu, *New Mexico Tech, USA*
Tianming Liu, *University of Georgia, USA*
Gary Livingston, *University of Massachusetts Lowell, USA*
Dang Long, *NYS Department of Health, Center for Medical Science, USA*
Hongtao Lu, *Shanghai JiaoTong University, China*
Wencong Lu, *Shanghai University, China*
Carol Lushbough, *Computer Science, University of South Dakota, USA*
Xiaotu Ma, *Shanghai Jiao Tong University, China*
Ian M. MacDonald, *College of Saint Rose, USA*
Anant Madabhushi, *Rutgers State University of New Jersey, USA*
Krzysztof Malczewski, *Poznan University of Technology, Poland*
Kezhi Mao, *Nanyang Technological University, Singapore*
Majid Masso, *George Mason University, USA*
Vasilis Megalooikonomou, Temple *University, USA*
Duoqian Miao, *Tongji University, China*
Armin R. Mikler, *University of North Texas, USA*
Nasir-ud-Din, *Institute of Molecular Sciences & Bioinformatics, Pakistan*
Jun Ni, *University of Iowa, USA*
Giuseppe Nicosia, *University of Catania, Italy*
Timothy O'Connor, *University of Cambridge, United Kingdom*
Manish Paliwal, *Southern Illinois University at Carbondale, USA*
Minseo Park, *University of Massachusetts, USA*
Vinhthuy Phan, *University of Memphis, USA*
Helen Pointkivska, *Kent State University, USA*
Aleksandar Poleksic, *University of Northern Iowa, USA*
Monika Ray, *Washington University in St. Louis, USA*
Tong Ruan, *East China University of Science and Technology, China*
Ann Rundell, *Weldon School of Biomedical Engineering, Purdue University, USA*
Meena K. Sakharkar, *Nanyang Technological University, Singapore*
Jacqueline Signorini, *University Paris 8, France*
Xiaofeng Song, *Nanjing University of Aeronautics and Astronautics, China*
Vojislav Stojkovic, *Morgan State University, USA*
Jonathan Sun , *University of Southern Mississippi, USA*
Xiao Sun, *Southeast University, China*
Jay Urbain, *Illinois Institute of Technology, USA*

Athanasios Vasilakos, *University of Peloponnese, Greece*
Thanos Vasilakos, *University of Western Macedonia, Greece*
Miguel A. Vegarodriguez, *University Extremadura, Spain*
Jason T. L. Wang, *New Jersey Institute of Technology, USA*
Liangjiang Wang, *Clemson University, USA*
Min Wang, *Hohai University, China*
Nan Wang, *University of Southern Mississippi, USA*
Ruizhi Wang, *Tongji University, China*
Yufeng Wang, *University of Texas at San Antonio, USA*
Yuhang Wang, *Southern Methodist University, USA*
Yu-Ping Wang, *University of Missouri-Kansas City, USA*
Zhe Wang, *East China University of Science and Technology, China*
Qishi Wu, *University of Memphis, USA*
Jiaoxiong Xia, Shanghai *Municipal Education Commission, China*
Min Xu, *University of Southern California, USA*
Weisheng Xu, *Tongji University, China*
Jianhua Xuan, *Virginia Tech, USA*
Kaiguo Yan, *Thomas Jefferson University, USA*
Jie Yang, *Shanghai JiaoTong University, China*
Laurence T. Yang, *Saint Francis Xavier University, Canada*
Yilong Yin, *Shandong University, China*
Jingkai Yu, *Wayne State University, USA*
Yan Yu, *Thomas Jefferson University Hospital, USA*
Erliang Zeng, *Department of Computer Science, University of Miami, USA*
Aidong Zhang, *State University of New York at Buffalo, USA*
Daoqiang Zhang, *NUAA, China*
Ji Zhang, *Shanghai Jiao Tong University School of Medicine, China*
Jun Zhang, *University of Kentucky, USA*
Min-Ling Zhang, *Hohai University, China*
Weixiong Zhang, *Washington University in St. Louis, USA*
Yanqing Zhang, *Georgia State University, USA*
Ying Zhao, *Tsinghua University, China*
Huiru Zheng, *University of Ulster, United Kingdom*
W. Jim Zheng, *Bioinformatics & Epidemiology Medical University of South Carolina, USA*
Yang Zhong, *Fudan University, China*
Jie Zhou, *Northern Illinois University, USA*
Leming Zhou, *University of Pittsburgh, USA*
Shuigeng Zhou, *Fudan University, China*
Xiaobo Zhou, *Cornell University, USA*
Zhi-Hua Zhou, *Nanjing University, China*
Hill Zhu, *Florida Atlantic University, USA*
Qingxin Zhu, *University of Electronic Science and Technology of China, China*
Shanfeng Zhu, *Fudan University, China*

Reviewers

Abdellali Kelil
Aidong Zhang
Albert C. S. Chung
Aleksandar Poleksic
Anant Madabhushi
Ann Rundell
Armin R. Mikler
Athanasios Vasilakos
Bernard Chen
Carol Lushbough
Changjin Hong
Chang-Shing Lee
Chang-Tsun Li
Christopher M. Frenz
Chun-Yuan Lin
Cornelia Caragea
Danail Bonchev
Dandan Li
Dang Long
Daoming Zhou
Daoqiang Zhang
David Keathly
Do Amaral
Doina Caragea
Dongsup Kim
Duoqian Miao
Elisa Ficarra
Erliang Zeng
Feipei Lai
Fu Chang
Gary Livingston
Giuseppe Nicosia
Haim Levkowitz
Helen Pointkivska
Hill Zhu
Hong Cai
Hongbo Zhou
Hongfang Liu
Hongtao Lu
Hongwei Huo
Huiru Zheng
Hyunsoo Kim
Ian M. MacDonald
Jacqueline Signorini
Jaeseung Jeong
Jaewoo Kang
Jake Y. Chen
Jason T. L. Wang
Jay Urbain
Jean X. Gao
Ji Zhang
Jia Li
Jianhua Xuan
Jianlin Cheng
Jianying Gu
Jiaoxiong Xia
Jie Yang
Jie Zhou
Jiejun Kong
Jin Ping
Jing Li
Jingchun Sun
Jingkai Yu
Jonathan Sun
Jun Liu
Jun Ni
Jun Zhang
Jun-tao Guo
Kaiguo Yan
Kang Li
Kevin Daimi
Kezhi Mao
Kreshna Gopal
Krzysztof Malczewski
Laurence Rodrigues
Laurence T. Yang
Lei Shi
Leming Zhou
Leng Han
Liangjiang Wang
Liangxiao Jiang
Liao Li
Mahua Bhattacharya
Majid Masso
Manish Paliwal
Maozu Guo
Mark Clement
Maya El Dayeh
Meena K. Sakharkar
Miaojun Han
Miguel A. Vegarodriguez
Min Wang
Min Xu
Min Zheng
Ming Li
Min-Ling Zhang
Minrui Fei
Minseo Park,
Mojie Duan
Monika Ray
Nan Wang
Nasir-ud-Din
Pan Du
Phuongan Dam
Pierre Beauseroy
Pingyong Li
Preetam Ghosh
Purushotham Bangalore
Qiang Cheng
Qingxin Zhu
Qingzhong Liu
Qishi Wu
Qiwen Dong
Qun Niu
Rui Camacho
Rui Chang
Rui Jiang
Rui Kuang
Ruizhi Wang
RyangGuk Kim
Sadaf Alam
Shanfeng Zhu
Shao Li
Shaozi Li
Shen Li
Shuigeng Zhou
Siling Wang
Simon Lin
Sun Kim
Sung-Bae Cho
Tamara Schneider
Thanos Vasilakos
Tianming Liu
Tianrui Li
Tim G. Lilburn
Timothy O'Connor
Tong Ruan
Vasilis Megalooikonomou
Vinhthuy Phan
Vojislav Stojkovic
W. Jim Zheng
Wangshu Zhang
Wei Kong
Weisheng Xu
Weixiong Zhang
Wencong Lu
Wen-Lian Hsu
Werner Dubitzky
Xiao Sun
Xiaobo Zhou
Xiaofeng Song
Xiaoli Li
Xiaolin Li
Xiaotu Ma
Xin Geng
Yan Yu
Yanfeng Cheng
Yang Gao
Yang Song

Yang Zhong
Yanqing Zhang
Ye Duan
Yidong Chen
Yilong Yin
Ying Zhao
Ying-Xin Li
Yisheng Ding
Yufei Huang
Yufeng Wang
Yuhang Wang
Yu-Ping Wang
Zhang Fengkai
Zhe Wang
Zhenran Jiang
Zhi-Hua Zhou
Zhiwei Cao
Zhu Li
Zoé Lacroix

Plenary

Keynote Abstracts

Locating a Few Useful Clusters in Large Biological Datasets: A Tale of Two Viewpoints

Joydeep Ghosh
Dept of Electrical and Computer Engineering
University of Texas at Austin
1 Univ. Station C0803, Austin, TX 78712, USA
Email: ghosh@ece.utexas.edu

I. Extended Abstract

A key application of clustering data obtained from sources such as microarrays, protein mass spectroscopy and phylogenetic profiles, is the detection of functionally related genes. Typically, only a small number of functionally related genes form meaningful groups, while the rest need to be ignored. Additional complications arise when there are several irrelevant experimental conditions, when the useful clusters occur at different resolutions/scales, and when genes participate in multiple biological processes, leading to multiple cluster memberships.

The pioneering work of Cheng and Church [1] focused a lot of attention to (bi)-clustering of microarray data. By identifying the most cohesive bi-clusters, i.e., subsets of genes and experiments over which there is coherent expression behavior rather than trying to cluster all the genes using all available features, the original biclustering approach was able to obtain more meaningful groups of genes. Since then a variety of improvements have been made using both biclustering oriented and subspace clustering oriented methods specialized for such data. However, these methods have provided limited progress, specially in the presence of a large number of irrelevant genes/conditions and due to the multi-membership problem.

We have recently developed two different approaches to simultaneously deal with all the complications of such microarray data analysis. These techniques are also relevant for certain other types of bioinformatics problems including spectroscopy data analysis. The first approach, Automated Hierarchical Density Shaving (Auto-HDS), is a framework that consists of a fast, hierarchical, density-based clustering algorithm and an unsupervised model selection strategy. Auto-HDS can automatically select clusters of different densities, present them in a compact hierarchy and rank individual clusters using an innovative stability criteria. This framework also provides a simple yet powerful 2-D visualization of the hierarchy of clusters that is useful for further interactive exploration. This approach is detailed in [2], which also present results on Gasch and Lee microarray datasets to show its effectiveness. Moreover, a public domain implementation called GeneDiver is available for practitioners to use via the internet, at http://www.ideal.ece.utexas.edu/ gunjan/genediver/.

An alternative approach to HDS is to specify a suitable data generation model, and then fit model parameters using the available data. For example, one can model observed data as a mixture of k multivariate Gassians, fit this mixture to data using the EM algorithm, and then get a k-cluster solution by assigning a given data point to the mixture component this is most likely to have produced it. How can one determine a suitable generative model for microarray data? A start can be made by noting that since clusters could exist in different subspaces of the feature space, a co-clustering algorithm that simultaneously clusters objects and features is often more suitable as compared to one that is restricted to traditional "one-sided" clustering. At the same time, it is important to discard irrelvant features and objects, and to allow arbitrary shaped and located co-clusters, including overlapping ones.

These objectives are achieved in two steps: In the first step, the Bregman Co-clustering algorithm [3] is adapted to automatically prune away non-informative data while simultaneously clustering both genes and conditions. This step results in finding co-clusters arranged in a grid structure, but only a predetermined number of rows and columns are assigned to the co-clusters. An agglomeration step then appropriately merges similar co-clusters to discover dense, arbitrarily positioned and even overlapping co-clusters.. The underlying probablity model related two a mixture model of exponential family distributions, where each mixture component is indexed by a pair of (row,col) cluster identity, and with a uniform background distribution to model the genes/conditions that are irrelevant.

The overall methodology has the following key features that distinguish it from existing co-clustering algorithms. (i) The ability to mine the most coherent co-clusters from large and noisy datasets; (ii) Detection of arbitrarily positioned and possibly overlapping co-clusters in a principled manner by iteratively minimizing a suitable cost function, (iii) Generalization to all Bregman divergences, including squared Euclidean distance, commonly used for clustering microarray data and I-divergence, commonly used for text data clustering; (iv) The ability to naturally deal with missing data values, without introducing random perturbations or bias in the data and (v) Efficient detection all the co-clusters simultaneously rather than sequentially, enabling scalability to large and high-dimensional datasets. For more details of the method and its empirically observed success, please refer to [4].

Given the two approaches that are philosophically and methodologically very different, the natural question is how does one determine the method that is most suitable? The

answer lies in the match of the dataset to certain modeling characteristics, showing that both approaches have complementary strengths, and that together, they form a formidable suite for catering to a wide range of bio-informatics related clustering requirements.

ACKNOWLEDGMENT

This article reflects joint work with Gunjan Gupta, Meghana Deodhar and Alex Liu, and was supported by NSF grant IIS-071342.

REFERENCES

[1] Y. Cheng and G. M. Church, "Biclustering of expression data," in *(ICMB)*, 2000, pp. 93–103.

[2] G. Gupta, A. Liu, and J. Ghosh, "Automated hierarchical density shaving: A robust, automated clustering and visualization framework for large biological datasets," *To appear: IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2009.

[3] A. Banerjee, I. Dhillon, J. Ghosh, S. Merugu, and D. Modha, "A generalized maximum entropy to bregman co-clustering and matrix approximation," *JMLR*, vol. 8, pp. 1919–1986, 2007.

[4] M. Deodhar, G. Gupta, J. Ghosh, H. Cho, and I. Dhillon, "A scalable framework for discovering coherent co-clusters in noisy data," in *(To appear: ICML)*, 2009.