Ion Măndoiu
Alexander Zelikovsky (Eds.)

Bioinformatics Research and Applications

LNBI 4463

Third International Symposium, ISBRA 2007
Atlanta, GA, USA, May 2007
Proceedings

Springer

Ion Măndoiu Alexander Zelikovsky (Eds.)

Bioinformatics Research and Applications

Third International Symposium, ISBRA 2007
Atlanta, GA, USA, May 7-10, 2007
Proceedings

Springer

Series Editors

Sorin Istrail, Brown University, Providence, RI, USA
Pavel Pevzner, University of California, San Diego, CA, USA
Michael Waterman, University of Southern California, Los Angeles, CA, USA

Volume Editors

Ion Măndoiu
University of Connecticut
Computer Science and Engineering Department
Storrs, CT 06269-2155, USA
E-mail: ion@engr.uconn.edu

Alexander Zelikovsky
Georgia State University
Department of Computer Science
Atlanta, GA 30303-3086, USA
E-mail: alexz@cs.gsu.edu

Library of Congress Control Number: 2007925296

CR Subject Classification (1998): J.3, H.2.8, F.1, F.2.2, G.3

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743
ISBN-10 3-540-72030-8 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-72030-0 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12050161 06/3180 5 4 3 2 1 0

Lecture Notes in Bioinformatics 4463

Edited by S. Istrail, P. Pevzner, and M. Waterman

Editorial Board: A. Apostolico S. Brunak M. Gelfand
T. Lengauer S. Miyano G. Myers M.-F. Sagot D. Sankoff
R. Shamir T. Speed M. Vingron W. Wong

Subseries of Lecture Notes in Computer Science

Preface

The 2007 International Symposium on Bioinformatics Research and Applications (ISBRA 2007), was held on May 7–10, 2007 at Georgia State University in Atlanta, Georgia. The ISBRA symposium provides a forum for the exchange of ideas and results among researchers, developers, and practitioners working on all aspects of bioinformatics and computational biology and their applications. ISBRA is the successor of the International Workshop on Bioinformatics Research and Applications (IWBRA), held May 22–25, 2005 in Atlanta, GA and May 28–31, 2006 in Reading, UK, in conjunction with the International Conference on Computational Science.

This year, 146 papers were submitted in response to the call for papers. Following a rigorous review process, the Program Committee selected 55 papers for publication in the proceedings and oral presentations at the symposium. The topics of selected papers covered a wide range of topics, including clustering and classification, gene expression analysis, gene networks, genome analysis, motif finding, pathways, protein structure prediction, protein domain interactions, phylogenetics, and software tools.

In addition to contributed talks, the ISBRA 2007 technical program included several tutorials and poster sessions, and features invited keynote talks by three distinguished speakers. Ming Li from University of Waterloo spoke on modern homology search, Laura L. Elnitski from the National Human Genome Research Institute spoke on bidirectional promoters in the human genome, and Mark Borodovsky from Georgia Institute of Technology spoke on *ab initio* gene finding.

We would like to thank all authors for submitting papers and presenting their work at the symposium. We would also like to thank the Program Committee members and external reviewers for volunteering their time and expertise to review and select symposium papers. We would like to extend special thanks to the Organizing, Publications, Finance, Publicity, and Posters Chairs, all of whom are listed on the following page, for their tremendous efforts in making ISBRA 2007 a great success. Last but not least, we would like to thank the General Chairs, Dan Gusfield and Yi Pan, for their leadership and guidance.

We hope you will find the technical program interesting and thought provoking. Enjoy!

May 2007

Ion Măndoiu
Alexander Zelikovsky

Conference Organization

General Chairs

Dan Gusfield, University of California, Davis, USA
Yi Pan, Georgia State University, USA

Program Chairs

Ion Măndoiu, University of Connecticut, USA
Alexander Zelikovsky, Georgia State University, USA

Organizing Chairs

Robert Harrison, Georgia State University, USA
Yanqing Zhang, Georgia State University, USA

Publications Chair

Raj Sunderraman, Georgia State University, USA

Finance Chairs

Anu Bourgeois, Georgia State University, USA
Akshaye Dhawan, Georgia State University, USA

Publicity Chairs

Kim King, Georgia State University, USA
Yingshu Li, Georgia State University, USA

Poster Chairs

Gulsah Altun, Georgia State University, USA
Dumitru Brinza, Georgia State University, USA

Program Committee

Gabriela Alexe
IBM Research, USA

Yonatan Aumann
Bar Ilan University, Israel

Danny Barash
Ben-Gurion University, Israel

Anne Bergeron
Université du Québec à Montréal, Canada

Piotr Berman
Penn State, USA

Paola Bonizzoni
Università degli Studi di Milano-Bicocca, Italy

Mark Borodovsky
Georgia Institute of Technology, USA

Anu Bourgeois
Georgia State University, USA

Daniel Brown
University of Waterloo, Canada

Liming Cai
University of Georgia, USA

Jake Yue Chen
IUPUI, USA

Luonan Chen
Osaka Sangyo University, Japan

Yixin Chen
Washington University, USA

Peter Damaschke
Chalmers University, Sweden

Bhaskar Dasgupta
University of Illinois at Chicago, USA

Sorin Draghici
Wayne State University, USA

Jun-Tao Guo
University of Georgia, USA

Robert Harrison
Georgia State University, USA

Jieyue He
Southeast University, China

Xiaohua Hu
Drexel University, USA

Hae-Jin Hu
Georgia State University, USA

Chun-Hsi Huang
University of Connecticut, USA

Ming-Yang Kao
Northwestern University, USA

Marek Karpinski
University of Bonn, Germany

John Kececioglu
University of Arizona, USA

Ed Keedwell
University of Exeter, UK

Guojun Li
University of Georgia, USA

Yiming Li
National Chiao Tung University, Taiwan

Yingshu Li
Georgia State University, USA

Jianzhong Li
Harbin Institute of Technology, China

Yixue Li
Shanghai Center for Bioinformation Technology, China

Guohui Lin
University of Alberta, Canada

Xiaohui Liu
Brunel University, UK

Shiyong Lu
Wayne State University, USA

Jingchu Luo
Peking University, China

Osamu Maruyama
Kyushu University, Japan

Kayvan Najarian
UNC at Charlotte, USA

Giri Narasimhan
Florida International University, USA

Craig Nelson
University of Connecticut, USA

Laxmi Parida
IBM T.J. Watson Research Center, USA

Mihai Pop
University of Maryland, USA

Alex Pothen
Old Dominion University, USA

Teresa Przytycka
NCBI, USA

Sven Rahmann
Universität Bielefeld, Germany

Sanguthevar Rajasekaran
University of Connecticut, USA

David Sankoff
University of Ottawa, Canada

Russell Schwartz
Carnegie Mellon University, USA

Hagit Shatkay
Queen's University, Canada

Jens Stoye
Universität Bielefeld, Germany

Raj Sunderraman
Georgia State University, USA

Sing-Hoi Sze
Texas A&M University, USA

El-Ghazali Talbi
Université des Sciences & Technologies de Lille, France

Esko Ukkonen
University of Helsinki, Finland

Ugo Vaccaro
Universitá di Salerno, Italy

Gwenn Volkert
Kent State University, USA

Jianxin Wang
Central South University, China

Zidong Wang
Brunel University, UK

Limsoon Wong
NUS, Singapore

Weili Wu
University of Texas at Dallas, USA

Fang Xiang Wu
University of Saskatchewan, Canada

External Reviewers

Lecture Notes in Bioinformatics

Vol. 4463: I. Măndoiu, A. Zelikovsky (Eds.), Bioinformatics Research and Applications. XV, 653 pages. 2007.

Vol. 4453: T. Speed, H. Huang (Eds.), Research in Computational Molecular Biology. XVI, 550 pages. 2007.

Vol. 4414: S. Hochreiter, R. Wagner (Eds.), Bioinformatics Research and Development. XVI, 482 pages. 2007.

Vol. 4366: K. Tuyls, R. Westra, Y. Saeys, A. Nowé (Eds.), Knowledge Discovery and Emergent Complexity in Bioinformatics. IX, 183 pages. 2007.

Vol. 4360: W. Dubitzky, A. Schuster, P.M.A. Sloot, M. Schroeder, M. Romberg (Eds.), Distributed, High-Performance and Grid Computing in Computational Biology. X, 192 pages. 2007.

Vol. 4345: N. Maglaveras, I. Chouvarda, V. Koutkias, R. Brause (Eds.), Biological and Medical Data Analysis. XIII, 496 pages. 2006.

Vol. 4316: M.M. Dalkilic, S. Kim, J. Yang (Eds.), Data Mining and Bioinformatics. VIII, 197 pages. 2006.

Vol. 4230: C. Priami, A. Ingólfsdóttir, B. Mishra, H.R. Nielson (Eds.), Transactions on Computational Systems Biology VII. VII, 185 pages. 2006.

Vol. 4220: C. Priami, G. Plotkin (Eds.), Transactions on Computational Systems Biology VI. VII, 247 pages. 2006.

Vol. 4216: M.R. Berthold, R.C. Glen, I. Fischer (Eds.), Computational Life Sciences II. XIII, 269 pages. 2006.

Vol. 4210: C. Priami (Ed.), Computational Methods in Systems Biology. X, 323 pages. 2006.

Vol. 4205: G. Bourque, N. El-Mabrouk (Eds.), Comparative Genomics. X, 231 pages. 2006.

Vol. 4175: P. Bücher, B.M.E. Moret (Eds.), Algorithms in Bioinformatics. XII, 402 pages. 2006.

Vol. 4146: J.C. Rajapakse, L. Wong, R. Acharya (Eds.), Pattern Recognition in Bioinformatics. XIV, 186 pages. 2006.

Vol. 4115: D.-S. Huang, K. Li, G.W. Irwin (Eds.), Computational Intelligence and Bioinformatics, Part III. XXI, 803 pages. 2006.

Vol. 4075: U. Leser, F. Naumann, B. Eckman (Eds.), Data Integration in the Life Sciences. XI, 298 pages. 2006.

Vol. 4070: C. Priami, X. Hu, Y. Pan, T.Y. Lin (Eds.), Transactions on Computational Systems Biology V. IX, 129 pages. 2006.

Vol. 4023: E. Eskin, T. Ideker, B. Raphael, C. Workman (Eds.), Systems Biology and Regulatory Genomics. X, 259 pages. 2007.

Vol. 3939: C. Priami, L. Cardelli, S. Emmott (Eds.), Transactions on Computational Systems Biology IV. VII, 141 pages. 2006.

Vol. 3916: J. Li, Q. Yang, A.-H. Tan (Eds.), Data Mining for Biomedical Applications. VIII, 155 pages. 2006.

Vol. 3909: A. Apostolico, C. Guerra, S. Istrail, P. Pevzner, M. Waterman (Eds.), Research in Computational Molecular Biology. XVII, 612 pages. 2006.

Vol. 3886: E.G. Bremer, J. Hakenberg, E.-H.(S.) Han, D. Berrar, W. Dubitzky (Eds.), Knowledge Discovery in Life Science Literature. XIV, 147 pages. 2006.

Vol. 3745: J.L. Oliveira, V. Maojo, F. Martín-Sánchez, A.S. Pereira (Eds.), Biological and Medical Data Analysis. XII, 422 pages. 2005.

Vol. 3737: C. Priami, E. Merelli, P. Gonzalez, A. Omicini (Eds.), Transactions on Computational Systems Biology III. VII, 169 pages. 2005.

Vol. 3695: M.R. Berthold, R.C. Glen, K. Diederichs, O. Kohlbacher, I. Fischer (Eds.), Computational Life Sciences. XI, 277 pages. 2005.

Vol. 3692: R. Casadio, G. Myers (Eds.), Algorithms in Bioinformatics. X, 436 pages. 2005.

Vol. 3680: C. Priami, A. Zelikovsky (Eds.), Transactions on Computational Systems Biology II. IX, 153 pages. 2005.

Vol. 3678: A. McLysaght, D.H. Huson (Eds.), Comparative Genomics. VIII, 167 pages. 2005.

Vol. 3615: B. Ludäscher, L. Raschid (Eds.), Data Integration in the Life Sciences. XII, 344 pages. 2005.

Vol. 3594: J.C. Setubal, S. Verjovski-Almeida (Eds.), Advances in Bioinformatics and Computational Biology. XIV, 258 pages. 2005.

Vol. 3500: S. Miyano, J. Mesirov, S. Kasif, S. Istrail, P. Pevzner, M. Waterman (Eds.), Research in Computational Molecular Biology. XVII, 632 pages. 2005.

Vol. 3388: J. Lagergren (Ed.), Comparative Genomics. VII, 133 pages. 2005.

Vol. 3380: C. Priami (Ed.), Transactions on Computational Systems Biology I. IX, 111 pages. 2005.

Vol. 3370: A. Konagaya, K. Satou (Eds.), Grid Computing in Life Science. X, 188 pages. 2005.

Vol. 3318: E. Eskin, C. Workman (Eds.), Regulatory Genomics. VII, 115 pages. 2005.

Vol. 3240: I. Jonassen, J. Kim (Eds.), Algorithms in Bioinformatics. IX, 476 pages. 2004.

Vol. 3082: V. Danos, V. Schachter (Eds.), Computational Methods in Systems Biology. IX, 280 pages. 2005.

Vol. 2994: E. Rahm (Ed.), Data Integration in the Life Sciences. X, 221 pages. 2004.

Vol. 2983: S. Istrail, M.S. Waterman, A. Clark (Eds.), Computational Methods for SNPs and Haplotype Inference. IX, 153 pages. 2004.

Vol. 2812: G. Benson, R.D.M. Page (Eds.), Algorithms in Bioinformatics. X, 528 pages. 2003.

Vol. 2666: C. Guerra, S. Istrail (Eds.), Mathematical Methods for Protein Structure Analysis and Design. XI, 157 pages. 2003.

Table of Contents

GFBA: A Biclustering Algorithm for Discovering Value-Coherent Biclusters*

Xubo Fei[1], Shiyong Lu[1], Horia F. Pop[2], and Lily R. Liang[3]

[1] Dept. of Computer Science, Wayne State University, USA
{xubo,shiyong}@wayne.edu
[2] Dept. of Computer Science, Babes-Bolyai University, Romania
hfpop@cs.ubbcluj.ro
[3] Dept. of Computer Science and IT, University of the District of Columbia, USA
lliang@udc.edu

Abstract. Clustering has been one of the most popular approaches used in gene expression data analysis. A clustering method is typically used to partition genes according to their similarity of expression under different conditions. However, it is often the case that some genes behave similarly only on a subset of conditions and their behavior is uncorrelated over the rest of the conditions. As traditional clustering methods will fail to identify such gene groups, the biclustering paradigm is introduced recently to overcome this limitation. In contrast to traditional clustering, a biclustering method produces biclusters, each of which identifies a set of genes and a set of conditions under which these genes behave similarly. The boundary of a bicluster is usually fuzzy in practice as genes and conditions can belong to multiple biclusters at the same time but with different membership degrees. However, to the best of our knowledge, a method that can discover fuzzy value-coherent biclusters is still missing. In this paper, (i) we propose a new fuzzy bicluster model for value-coherent biclusters; (ii) based on this model, we define an objective function whose minimum will characterize good fuzzy value-coherent biclusters; and (iii) we propose a genetic algorithm based method, Genetic Fuzzy Biclustering Algorithm (GFBA), to identify fuzzy value-coherent biclusters. Our experiments show that GFBA is very efficient in converging to the global optimum.

1 Introduction

Clustering has been one of the most popular approaches used in gene expression data analysis. It is used to group genes according to their expression under multiple conditions or to group conditions based on the expression of a number of genes. When a clustering method is used for grouping genes, it typically partitions genes according to their similarity of expression under all conditions. However, it is often the case that some genes behave similarly only on a subset of conditions and their behavior is uncorrelated over the rest of the conditions. Therefore, traditional clustering methods will fail to identify such gene groups.

* This work was partially supported by the Agricultural Experiment Station at the University of the District of Columbia (Project No.: DC-0LIANG; Accession No.: 0203877).

I. Măndoiu and A. Zelikovsky (Eds.): ISBRA 2007, LNBI 4463, pp. 1–12, 2007.