

36th Symposium on the Interface: Computing Science and Statistics 2004

Computational Biology and Bioinformatics

Computing Science and Statistics Volume 36

**Baltimore, Maryland, USA
26-29 May 2004**

Volume 1 of 2

ISBN: 978-1-61567-070-3

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2004) by the Interface Foundation of North America
All rights reserved.

Printed by Curran Associates, Inc. (2009)

For permission requests, please contact the Interface Foundation of North America
at the address below.

Interface Foundation of North America
PO Box 7460
Fairfax Station, VA 22039

Phone: (703) 993-1212
Fax: (703) 993-1700

interface@galaxy.gmu.edu

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

VOLUME 1

Interface 2004 Schedule (with titles and speakers).....	1
Interface 2004 Abstracts.....	18
Short Course I: Random Forests.....	91
<i>Leo Breiman and Adele Cutler</i>	
Short Course II: Gene Expression Analysis.....	159
<i>Rafael A. Irizarry</i>	
Comparison of Classification Techniques in Bioinformatics.....	160
<i>Rashpal Ahluwalia and Sundar Chidambaram</i>	
Visualizing Patients Treated with Three-Dimensional Computed Tomography-Guided Brachytherapy.....	164
<i>Faleh Alshameri and Jee Vang</i>	
Extending the Loop Design for Microarray Experiments.....	186
<i>Naomi S. Altman</i>	
Mixed Effects Model for Assessing RNA Degradation in Affymetrix GeneChip.....	196
Experiments <i>Kellie J. Archer, Suresh E. Joel, and Viswanathan Ramakrishnan</i>	
Modeling Dinucleotide Density Fluctuations in Genome Sequences.....	197
<i>R. H. Baran</i>	
Mining Distance-Based Outliers in Near Linear Time.....	212
<i>Stephen D. Bay and Mark Schwabacher</i>	
Five Hierarchical Levels of Sequence-Structure Correlation in Proteins.....	229
<i>Christopher Bystroff</i>	
Evaluating Natural Language Processing Applications Applied to Outbreak and Disease Surveillance.....	243
<i>Wendy W. Chapman, John N. Dowling, Oleg Ivanov, Per H. Gesteland, Robert T. Olszewski, Jeremy U. Espino, and Michael M. Wagner</i>	
Learning Imbalanced Data with Random Forests.....	263
<i>Chao Chen, Andy Liaw, and Leo Breiman</i>	
Limitations of Statistical Learning from Gene Expression Data.....	266
<i>Tianjiao Chu</i>	

The Restricted Partition Method for Detecting Epistatic Interactions Contributing to a Quantitative Trait.....	286
<i>Robert Culverhouse, Tsvika Klein, Mary Relling, and William Shannon</i>	
The MiTAP System for Monitoring Reports of Disease Outbreak.....	296
<i>Laurie E. Damianos, Guido Zarrella, and Lynette Hirschman</i>	
Cluster Substance Identification via Conditional Entropy Calculations.....	308
<i>James C. Diggans and Jeffrey L. Solka</i>	
A Wavelet-Based Statistical Analysis of fMRI Data: I. Motivation and Data Distribution Modeling.....	325
<i>Ivo D. Dinov, John W. Boscardin, Michael S. Mega, Elizabeth L. Sowell, and Arthur W. Toga</i>	
On the Least Median Square Problem.....	351
<i>Jeff Erickson, Sarel Har-Peled, and David Mount</i>	
Jointly Optimizing Model Complexity and Data-Processing Parameters with Mixed-Input SPSA.....	354
<i>Jim Garrett</i>	
Gene Expression Comparisons for Class Prediction in Cancer Studies.....	369
<i>Donald Geman, Christian d'Avignon, Daniel Q. Naiman, Raimond L. Winslow, and Arnaud Zeboulon</i>	
Estimating the Parameters of Infinite Scale Mixtures of Normals.....	384
<i>Hasan Hamdan and John P. Nolan</i>	
Yxilon-Designing the Next Generation, Vertically Integrable Statistical Software Environment.....	400
<i>Wolfgang Härdle, Sigbert Klinke, and Uwe Ziegenhagen</i>	
Computation of the k^{th} Nearest Neighbor Estimate of Entropy of Molecules Using Parallel Processing.....	416
<i>E. James Harner, Jun Tan, Shengqiao Li, and Harshinder Singh</i>	
Assessing Survival Forests for Prognosis Based on Gene Profiles.....	425
<i>Thu M. Hoàng and Van L. Parsons</i>	
Application of the Random Forest Classification Algorithm to a SELDI-TOF Proteomics Study in the Setting of a Cancer Prevention Trial.....	433
<i>Grant Izmirlian</i>	

Characterization and Re-Annotation of Common Genes Found in 35 Complete Chloroplast Genomes.....	455
<i>Beatrice Kilel</i>	

Structural Analysis of Network Traffic Flows.....	464
<i>Eric Kolaczyk</i>	

Assessment of the Relative Therapeutic Effect in Small Groups at Several Time Points: Comparison of Mucosal and Subcutaneous Peptide Vaccines in Rhesus Macaques Exposed to SHIV.....	472
<i>V.A. Kuznetsov, V.S. Stepanov, J.A. Berzofsky, and I.M. Belyakov</i>	

Subsampling Model Selection in Neural Networks for Nonlinear Time Series Analysis.....	491
<i>Michelle La Rocca and Cira Perna</i>	

Intersection Graphs for Text Analysis.....	507
<i>Elizabeth Leeds and David J. Marchette</i>	

Shakespeare: A Combinatoric to Gene Network Modularization.....	517
<i>Nicholas Lewin-Koh and Christopher Taylor</i>	

VOLUME 2

User Profiling in Window Title and Process Table.....	530
<i>Chien-Chih Lin, Eun Young Noh, Youngping Yan, and Edward Wegman</i>	

The Volume-of-Tube Formula: Computational Methods and Statistical Applications..	547
<i>Catherine Loader</i>	

SVD-Based Functional ANOVA for Measurement Evaluation of MALDI-TOF Mass Spectrometry of Polymers.....	562
<i>Z. Q. John Lu</i>	

Monte Carlo Analysis of Univariate Statistical Outlier Techniques.....	566
<i>Mark W. Lukens</i>	

Automated Phenotypic Networks for the Integration of Heterogeneous Databases.....	573
<i>Yves A. Lussier and Xiaoyan Wang</i>	

Signal Conditioning and Filtering of SELDI Mass Spectrometry Time Series.....	582
<i>Dariya I. Malyarenko, William E. Cooke, Eugene R. Tracy, Haijian Chen, O. John Semmes, Maciek Sasianowski, Michael W. Trosset, and Dennis M. Manos</i>	

Confidence-Based Cost-Sensitive Classification Decisions.....	589
<i>Dragos D. Margineantu</i>	

Classification and Clustering Using Weighted Text Proximity Matrices.....	600
<i>Wendy L. Martinez, Angel R. Martinez and Edward J. Wegman</i>	
Polyoptimizing Genetic Algorithm for Feature Subset Selection.....	612
<i>Ewy Mathe and John Grefenstette</i>	
Identifying Differentially Expressed Proteins in 2-D DIGE Experiments.....	622
<i>Yan Ma and E. James Harner</i>	
Multivariate Density Estimation for Massive Datasets via Sequential Convex Hull Peeling.....	632
<i>James P. McDermott and Dennis K. J. Lin</i>	
Supervised Learning Methods for Gene-Expression Data.....	659
<i>G. J. McLachlan, C. Ambroise, L. Ben-Tovim Jones, and X. Zhu</i>	
XML-Based Applications in Statistical Analysis.....	679
<i>Yuichi Mori, Tomokazu Fujino, Yoshiro Yamamoto, Takafumi Kubota, and Tomoyuki Tarumi</i>	
A Comparison of Sequential and False Discovery Rate Algorithms: Computational Experiments for Exploratory DNA Microarray Studies.....	691
<i>Danh V. Nguyen</i>	
Streaming Graphics.....	706
<i>Andrew Norton and Leland Wilkinson</i>	
Gene-Gene and Gene-Environment Interactions and Genetic Case-Control Association Studies.....	712
<i>Jurg Ott and Josephine Hoh</i>	
Wavelet and SiZer Analysis of Internet Traffic Data.....	716
<i>Cheolwoo Park, Fred Godtliebsen, Felix Hernandez-Campos, J. S. Marron, Vitaliana Rondonotti, F. Donelson Smith, Stilian Stoev, and Murad Taqqu</i>	
An Efficient Max-Dependency Algorithm for Gene Selection.....	729
<i>Hanchuan Peng and Fuhui Long</i>	
Evolving Classifiers for Knowledge Discovery in Medical and Biological Databases..	743
<i>Michael R. Peterson, Travis E. Doom, and Michael L. Raymer</i>	
DNA Microbial and Viral Identification Using Ultra Specific Probes “Blind” to Host Background DNA.....	762
<i>Catherine Putonti, George Fox, Richard C. Willson, B. Montgomery Pettitt, and Yuriy Fofanov</i>	

Wavelet Domain Linear Inversion via the LASSO.....	772
<i>Leming Qu and Partha Routh</i>	
Computational Geometry, Data Depth, and Robust Statistics.....	784
<i>Eynat Rafalin and Diane Souvaine</i>	
Noncentral Generalized F Distributions with Applications to Joint Outlier Detection...	801
<i>Donald E. Ramirez</i>	
Actor Allegiance and Block Model Strength.....	826
<i>John Rigsby and Jeffrey L. Solka</i>	
Performance of the False Discovery Rate for Small Sets of cDNA Microarrays.....	832
<i>Simon Rosenfeld</i>	
Fitting Large-Scale Spatial Models with Applications to Microarray Data Analysis....	869
<i>Stephan R. Sain and Reinhard Furrer</i>	
Bayesian Hierarchical Model of the Browsing Behavior of World Wide Web Users...	884
<i>Juana Sanchez and Ching-Ti Liu</i>	
Alternatives to Mixture Modeling in High Dimensions.....	897
<i>David W. Scott</i>	
Visual Data Mining of RNA Secondary Structure and Folding Pathways as Determined by the Massive Parallel Genetic Algorithm.....	916
<i>Bruce A. Shapiro and Wojciech Kasprzak</i>	
Parallelizing the Computation of Spatial Covariance in Large Spatial Datasets.....	935
<i>James A. Shine</i>	
An Efficient Algorithm for Simulating Coalescence with Recombination.....	941
<i>Katy L. Simonsen, Dan A. Noland, and Chinh Le</i>	
Model-Based Clustering with an Adaptive Mixtures Smart Start.....	946
<i>Jeffrey L. Solka and Wendy L. Martinez</i>	
Identifying Cross Corpora Document Associations via Minimal Spanning Trees.....	952
<i>Jeffrey L. Solka, Ivory C. Bryant, and Edward J. Wegman</i>	
The Analysis of Biomedical Data-Caveats and Challenges.....	962
<i>Ray L. Somorjai</i>	
Cramér-Rao Bounds and Monte Carlo Calculation of the Fisher Information Matrix in Difficult Problems.....	975
<i>James C. Spall</i>	

On the Construction of Discriminant Coordinates from Dissimilarity Data.....	990
<i>Michael W. Trosset</i>	
Privacy-Preserving k-Means Clustering over Vertically Partitioned Data.....	998
<i>Jaideep Vaidya and Chris Clifton</i>	
Mining Concepts-Drifting Data Streams.....	1008
<i>Haixum Wang</i>	
Visual Analytics for Streaming Internet Traffic.....	1023
<i>Edward J. Wegman</i>	
Performance Metrics for Group-Detection Algorithms.....	1032
<i>J. V. White, S. Steingold, and C. G. Fournelle</i>	
A Two-Stage Nearest-Neighbor Classifier with Application to Microbial Source Tracking.....	1047
<i>Jayson D. Wilbur</i>	
Indexing Continual Range Queries for Efficient Steam Processing.....	1050
<i>Kun-Lung Wu, Shyh-Kwei Chen, and Philip S. Yu</i>	
Visual Analytics for Dynamically Conditioned Choropleth Maps: QQplots, Scatterplots, Smoothes, and Two-Way Tables.....	1066
<i>Chunling Zhang, Yaru Li, and Daniel Carr</i>	
Index.....	1080