BIOKDD04: Workshop on Data Mining in Bioinformatics
August 22nd, 2004
Seattle, WA, USA

in conjunction with

10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

Mohammed J. Zaki
Computer Science Department
Rensselaer Polytechnic Institute
Troy, NY 12180, USA
zaki@cs.rpi.edu

Shinichi Morishita
Department of Computational Biology
University of Tokyo
Kashiwa City, Chiba, Japan
moris@k.u-tokyo.ac.jp

Isidore Rigoutsos, PhD
Manager, Bioinformatics & Pattern Discovery Group
IBM Thomas J Watson Research Center
Yorktown Heights, NY 10598
rigoutso@us.ibm.com

Opening Remarks

Bioinformatics is the science of managing, mining, and interpreting information from biological sequences and structures. Genome sequencing projects have contributed to an exponential growth in complete and partial sequence databases. The structural genomics initiative aims to catalog the structure-function information for proteins. Advances in technology such as microarrays have launched the subfield of genomics and proteomics to study the genes, proteins, and the regulatory gene expression circuitry inside the cell. What characterizes the state of the field is the flood of data that exists today or that is anticipated in the future; data that needs to be mined to help unlock the secrets of the cell.

While tremendous progress has been made over the years, many of the fundamental problems in bioinformatics, such as protein structure prediction or gene finding, are still open. Data mining will play a fundamental role in understanding gene expression, drug design and other emerging problems in genomics and proteomics. Furthermore, text mining will be fundamental in extracting knowledge from the growing literature in bioinformatics.

The goal of this workshop is to encourage KDD researchers to take on the numerous challenges that Bioinformatics offers. The workshop features an invited talks from noted expert in the field, and the latest data mining research in bioinformatics. We encouraged papers that propose novel data mining techniques for tasks such as:

- Gene expression analysis
- Protein/RNA structure prediction
- Phylogenetics
- Sequence and structural motifs
- Genomics and Proteomics
- Gene finding
- Drug design
- RNAi and microRNA Analysis
- Text mining in bioinformatics
- Modeling of biochemical pathways

These proceedings contain 10 papers (6 long and 4 short), out of 26 submissions that were accepted for presentation at the workshop. Each paper was reviewed by three members of the program committee. Along with a keynote talk, we were able to assemble a very exciting program.

We would like to thank all the authors, invited speaker, and attendees for contributing to the success of the workshop. Special thanks are due to the program committee for help in reviewing the submissions.

This workshop follows the previous three highly successful workshops: BIOKDD03, held in Washington, DC; BIOKDD02, held in Edmonton, Canada; and BIOKDD01 held in San Francisco, CA. We expect BIOKDD04 to be equally successful.

Workshop Co-Chairs

- Mohammed J. Zaki, Rensselaer Polytechnic Institute
- Shinichi Morishita, University of Tokyo
- Isidore Rigoutsos, IBM T.J. Watson Research Center

Program Committee

• **Srinivas Aluru**, Iowa State U., USA • **Alberto Apostolico**, Prudue U., USA • **Tatsuya Akutsu**, Kyoto U., Japan • **Charles Elkan**, UC San Diego, USA • **Jayant Haritsa**, Indian Inst. of Science, India • **Hasan Jamil**, Wayne State U., USA • **Andreas Karwath**, U. Freiburg, Germany • **George Karypis**, U. Minnesota, USA • **Ross D. King**, U. of Wales, UK • **Jinyan Li**, Inst. for Infocomm Research, Singapore • **Lance A. Liotta**, NIH/NCI, USA • **Ambuj Singh**, UC Santa Barbara, USA • **David Page**, U. Wisconsin, USA • **Srinivasan Parthasarathy**, Ohio State U., USA • **Jignesh M. Patel**, U. Michigan, USA • **Daniel E. Platt**, IBM TJ Watson, USA • **Luc De Raedt**, U. Freiburg, Germany • **Tobias Scheffer**, Humboldt U., Germany • **Karlton Sequeira**, RPI, USA • **Hannu Toivonen**, U. Helsinki, Finland • **Jason Wang**, NJIT, USA • **Wei Wang**, UNC Chapel-hill, USA • **Jiong Yang**, Case Western Reserve U., USA • **Aidong Zhang**, U. Buffalo, USA

Workshop Program & Table of Contents