BIOKDD '07: Workshop on Data Mining in Bioinformatics August 12th, 2007 San Jose, CA, USA

in conjunction with
13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

Jake Y. Chen
School of Informatics
Indiana University
Indianapolis, IN 46202
jakechen@iupui.edu

Stefano Lonardi
Dept. of Computer Science and Eng.
University of California
Riverside, CA 92521
stelo@cs.ucr.edu

Mohammed Zaki
Department of Computer Science
Rensselaer Polytechnic Institute
Troy, NY 12180-3590
zaki@cs.rpi.edu

REMARKS

Bioinformatics is the science of managing, mining, and interpreting information from biological processes. Various genome projects have contributed to an exponential growth in DNA and protein sequence databases. Advances in high-throughput technology such as microarrays and mass spectrometry have further created the fields of functional genomics and proteomics, in which one can monitor quantitatively the presence of multiple genes, proteins, metabolites, and compounds in a given biological state. The ongoing influx of these data, the presence of biological answers to data observed despite noises, and the gap between data collection and knowledge curation have collectively created new and exciting opportunities for data mining researchers in the post-genome era.

While tremendous progress has been made over the years, many of the fundamental problems in bioinformatics, such as protein structure prediction, gene-environment interaction, and molecular pathway mapping, are still open. Data mining will play essential roles in understanding these fundamental problems and developing novel therapeutic/diagnostic solutions in post-genome medicine.

Data mining approaches seem ideally suited for bioinformatics, since the field is awash with data from high-throughput experimental instruments. The extensive databases of biological information available create both challenges and opportunities for developing novel knowledge discovery and data mining methods. To provide avenues to data mining researchers active in bioinformatics, we have been organizing the Workshops on Data Mining in Bioinformatics (BIOKDD), held annually in conjunction with the ACM SIGKDD Conference in 2001-2006. This is the 7th year for the workshop.

The goal of this year's workshop call for papers (CFP) was to encourage KDD researchers to take on the numerous research challenges that bioinformatics offers. In our CFP, we encouraged paper submissions that present novel data mining techniques in the following sample topics:

- Phylogenetics and comparative Genomics
- DNA microarray data analysis
- RNAi and microRNA Analysis
- Protein/RNA structure prediction
- Sequence and structural motif finding
- Modeling of biological networks and pathways
- Statistical learning methods in bioinformatics

- Computational proteomics
- Computational biomarker discoveries
- Computational drug discoveries
- Biomedical text mining
- Biological data management techniques
- Semantic webs and ontology-driven biological data integration methods

PROGRAM

The workshop is a full day event in conjunction with the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, CA, August 12-15, 2007. The workshop was accepted in the conference program after the SIGKDD conference organization committee reviewed the competitive proposal submitted by the workshop co-chairs. To promote this year's program, we established an Internet web site at http://bio.informatics.iupui.edu/biokdd07.

This year, we accepted 10 papers out of 24 submissions into the workshop program and proceedings due to the exceptionally high quality of the submissions. Among these papers, 7 of the papers are accepted as full presentations (30 minutes each) and 3 of the papers are accepted as short presentations (20 minutes each). Each paper was peer reviewed by three members of the program committee and papers with declared conflict of interest were reviewed blindly to ensure impartiality. All papers, whether accepted or rejected, were given detailed review forms as a feedback.

In closing, we want to thank Atul Butte, M.D., Ph.D. who agreed to give the keynote talk for this year's program. Dr. Butte is an Assistant Professor in Medicine (Medical Informatics) and Pediatrics at the Stanford University School of Medicine and the Lucile Packard Children's Hospital. His talk is entitled "Exploring Genomic Medicine Using Integrative Biology".

WORKSHOP CO-CHAIRS

- Jake Y. Chen, Indiana University – Purdue University, Indianapolis
- Stefano Lonardi, University of California, Riverside
- Mohammed J. Zaki, Rensselaer Polytechnic Institute (General Chair)

PROGRAM COMMITTEE

Amandeep Sidhu (Curtin University, Australia), Eamonn Keogh (University of California, Riverside), Daisuke Kihara (Purdue University), Giuseppe Lancia (University of Udine, Italy), Guojun Li (ShanDong University, China), Haixu Tang (Indiana University), Huanmei Wu (IUPUI), Isidore Rigoutsos (IBM T. J. Watson Research Center), Jason Wang (New Jersey Institute of Technology), Jie Zheng (NCBI, USA), Jignesh M. Patel (University of Michigan), Knut Reinert (Freie Universitt Berlin, Germany), Li Liao (University of Delaware), Luke Huan (University of Kansas), Fenglou Mao (University of Georgia), Muhammad Abulaish (Jamia Millia Islamia, India), Natasa Przulj (University of California, Irvine), Pan Du (Northwestern University), Phoebe Chen (Deakin University, Australia), Rahul Singh (San Francisco State University), Richard Scheuermann (University of Texas Southwestern), Simon Lin (Northwestern University), Xiang Zhang (Purdue University), Teresa Przytycka (NCBI/NLM, USA), Tony Hu (Drexel University), Xiaoyan Zhu (Tsinghua University, China), Yi Pan (Georgia State University), Yu-Ping Wang (University of Missouri)

ACKNOWLEDGEMENTS

We would like to thank all the program committee members, contributing authors, invited speaker, and attendees for contributing to the success of the workshop. Special thanks are also extended to the SIGKDD '07 conference organizing committee, particularly Qiang Yang, for coordinating with us to put together the excellent workshop program on schedule.

WORKSHOP SCHEDULE AND INDEX TO PROCEEDING