

Guest Editorial for Special Section on BOKDD2013

Gaurav Pandey and Huzefa Rangwala

OVER the last few decades, huge amounts and large variety of data have been generated to study complex biological processes and diseases. These “big data” arise from efforts involving sequencing, imaging, genomics, proteomics and other biomedical technologies. The scale and complexity of these data sets have shifted the focus to their management, mining and integration, thus necessitating the development of advanced computational methods that can narrow the otherwise broad space of potentially valid biological and medical hypotheses. These methods employ concepts from diverse areas, such as data mining, statistics, computer science and physics, and are applied to a wide variety of data, often in an integrative manner.

The International Workshop on Data Mining in Bioinformatics (BOKDD), held in conjunction with the ACM Conference on Knowledge Discovery and Data Mining (KDD) for 12 years, has successfully established a tradition of providing a platform for the presentation and discussion of advances in data mining techniques that primarily target biological data. In 2013, the workshop was held in Chicago, and BOKDD continued the tradition of bringing together data mining researchers and life scientists, emphasizing novel problems with various types of biological data. This special section features extended versions of four papers that were presented at BOKDD.

In “Network-Based Drug Target Interaction Prediction with Probabilistic Soft Logic”, Shobeir Fakhraei, Bert Huang, Louiqa Raschid, and Lise Getoor propose a probabilistic soft logic (PSL)-based method to identify novel drug targets, which in turn provides a way to reposition drugs. By representing drug-target relationships as a bipartite graph and specifying PSL-derived rules based on this graph, they infer novel links connecting drugs and targets. By extensive experiments on multiple drug databases, they demonstrate that this approach can infer novel drug targets accurately and efficiently. They also conduct literature-based validation of several high-confidence predictions made by their method.

In “Reliable Radiation Hybrid Maps: An Efficient Scalable Clustering-based Approach”, Raed I. Seetan, Anne M. Denton, Omar Al-Azzam, Ajay Kumar, M. Javed Iqbal, and Shahryar F. Kianian present a scalable clustering-based approach for mapping markers from radiation hybrid mapping (RHM) experiments. The clustering step in this approach enables the efficient mapping of markers by imposing a divide-and-conquer structure on this otherwise computationally very expensive problem. Experiments on RHM data of the human genome illustrate that the proposed approach have lower computational complexity and produces better framework maps than state of the art tools for this task.

In “LateBiclustering: Efficient Heuristic Algorithm for Time-Lagged Bicluster Identification”, Joana P. Gonçalves and Sara C. Madeira present an efficient biclustering algorithms to identify time-lagged local coherence patterns from time-series data sets. Although this problem is NP-hard in its native form, the authors present several suffix tree-based algorithms to identify maximal time-lagged biclusters in linear time. By applying these algorithms to time series gene expression data, they demonstrate that the discovered biclusters consists of functionally-related genes and thus biologically significant.

Finally, in “Discovery of Spatially Cohesive Itemsets in Three-Dimensional Protein Structures”, Cheng Zhou, Pieter Meysman, Boris Cule, Kris Laukens, and Bart Goethals study the problem of efficiently discovering frequent cohesive substructures (itemsets) from three-dimensional protein structures. For this, they define the support of a substructure based on the spatial proximity among the included amino acids. By using this support function within an itemset discovery algorithm, they discover several frequent substructures among proteins belonging to specific functional families. These substructures are shown to represent biologically important entities like enzyme active sites.

As guest editors of this special section, we would like to thank the contributing authors, BOKDD program committee, the reviewers who reviewed the papers for this special section, and the TCBB editorial staff for their invaluable contributions.

Gaurav Pandey
Huzefa Rangwala
Guest Editors

- G. Pandey is with the Department of Genetics and Genomic Science and the Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029. E-mail: gaurav.pandey@mssm.edu.
- H. Rangwala is with the Department of Computer Science, George Mason University, Fairfax, VA 22030. E-mail: rangwala@cs.umn.edu.

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TCBB.2014.2348731



Gaurav Pandey received the PhD degree in computer science and engineering from the University of Minnesota, Twin Cities, in 2010, and subsequently completed the postdoctoral fellowship from the University of California, Berkeley. He is an assistant professor in the Department of Genetics and Genomic Sciences and the Icahn Institute for Genomics and Multiscale Biology at the Icahn School of Medicine at Mount Sinai, New York. His primary fields of interest are computational biology, genomics and large-scale

data analysis and mining, and he has published extensively in these areas. In 2011, he received a Certificate of Recognition by ACM SIGKDD for his dissertation research.



Huzefa Rangwala received the PhD degree in computer science from the University of Minnesota in 2008. He is an associate professor in the Department of Computer Science, George Mason University, VA. His research interests include bioinformatics, machine learning, and high-performance computing. He received the US National Science Foundation (NSF) Early Faculty Career Award in 2013, the 2014 George Mason Teaching Excellence Award, the 2013 Volgenau Outstanding Teaching Faculty Award, 2012 Computer Science Department Outstanding Teaching Faculty Award, and 2011 Computer Science Department Outstanding Junior Researcher Award.

▷ **For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.**