Introduction to the Minitrack on
Data Analysis and Visualization in Biomedical Informatics

Hesham H. Ali and Kathryn M. Dempsey

College of Information Science and Technology, University of Nebraska at Omaha, Omaha, NE 68182

The increasing availability of "big" biological and medical data continues to constitute significant challenges as well as opportunities in biomedical research. New technologies, such as next-generation sequencing, and new health-related policies, like the ones associated with electronic medical records, promise to produce even more data in the near future. The main research question in biomedical sciences has been how to extract useful knowledge from such massive raw data. The development of innovative tools to integrate, analyze and mine such data sources is a key step towards achieving large impact levels. In addition, advanced tools for visualizing biomedical data at various processing stages are critical in maximizing the value of its utilization. As a result, attention has been shifting recently for a focus on data generation technologies to data analysis and data visualization tools. Such tools are critical in taking full advantage of the public and private data currently available to all biomedical researchers.

The focus of data analysis and data visualization in biomedical research highlights the current state of research in the key biomedical research areas such as bioinformatics, medical informatics, public health informatics and biomedical imaging. Main topics to be covered in this mini-track include development of algorithms and tools aimed to solve the vast spectrum of challenging data utilization problems appearing in health care and in various areas related to biomedical research, particularly issues related to the ability to utilize Biological and Clinical databases.

The first paper of the mini-track, "Learning Predictive Models from Integrated Healthcare Data: Extending Pattern-based and Generative Models to Capture Both Temporal and Cross-Attribute Dependencies", investigates how to properly model attributes in temporal datasets, specifically those that are sparse. The goal of the study is to create a model that is able to anticipate a health condition or event based on a time sequence of records. The authors propose that sparse time sequence data can be used to anticipate health conditions; however no such methods currently exist that are able to accurately pinpoint these events. The proposed model captures time- and attribute- related dependencies using

Hidden Markov Models and pattern-based methods.

Next in the mini-track is a paper entitled "Application of a Hybrid Text Mining Approach to the Study of Suicidal Behavior in a Large Population," wherein the authors address note that suicide prevention is a high-priority concern for veterans especially due to a known elevated risk in this group. Using a large cohort of 250,000 patients, the authors a hybrid system that combines a search engine and a natural language feature extraction and classification system to estimate annual incidence of suicide attempts. Further, they study the speculated link between suicide risk and the impact of adverse childhood experiences. The authors find a positive association between suicide attempt incidence and a history of childhood abuse, neglect or family dysfunction, and that risk increases when multiple adverse events are reported.

The third article in the mini-track is titled "SNP-SNP Interaction using Gauss Chaotic Map Particle Swarm Optimization to Detect Susceptibility of Breast Cancer." In this work, the authors describe a Particle Swarm Optimization scheme, referenced as "Gauss PSO", in the context of evaluating SNP combinations. Comparison of their method to a traditional PSO scheme reveals that certain combinations of SNPs offer a protective effect against breast cancer. The performance of their novel method (Gauss PSO) was found to be an improvement on others, such as the PSO approach, on 2- to 6- order combinations of SNPs.

The continuous growth of biomedical data and the need to learn as much as possible from it represent one of the most challenging research area of our time. In addition, the process of penetrating IT advancements in the medical profession continues to be another challenging problem. There are many components that are critical to address this challenge but we believe that the development of innovative tools to integrate, analyze and visualize such data is essential to achieve large impact levels. We hope that the articles included in this mini-track provide another step towards achieving this objective.

IEEE
computer society