

COMPUTER SCIENCE
GRADUATE STUDENT SEMINAR

Data Warehousing and Ensemble Learning of Omics Data

Xiaobo Sun
Emory University

Abstract: The development and application of high-throughput genomics technologies has resulted in massive quantities of diverse omics data that continue to accumulate rapidly. These rich datasets offer unprecedented and exciting opportunities to address long standing questions in biomedical research. However, our ability to explore and query the content of diverse omics data is very limited. Existing dataset search tools rely almost exclusively on the metadata. A text-based query for gene name(s) does not work well on datasets where the vast majority of their content is numeric. To overcome this barrier, we have developed Omicseq, a novel web-based platform that facilitates the easy interrogation of omics datasets holistically, beyond just metadata to improve findability. The core component of Omicseq is trackRank, a novel algorithm for ranking omics datasets that fully uses the numerical content of the dataset to determine relevance to the query entity. The Omicseq system is supported by a scalable and elastic, NoSQL database that hosts a large collection of processed omics datasets. In the front end, a simple, web-based interface allows users to enter queries and instantly receive search results as a list of ranked datasets deemed to be the most relevant. Omicseq is freely available at <http://www.omicseq.org>.

Friday, April 20, 2018, 1:00 pm
Room GCR311 of Department of Biostatistics

MATHEMATICS AND COMPUTER SCIENCE
EMORY UNIVERSITY