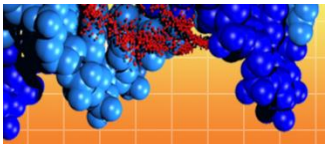

CENTER FOR BIOINFORMATICS & COMPUTATIONAL BIOLOGY

BIOINFORMATICS 2016 FALL SEMINAR SERIES

Hosted by: Department of Computer and Information Sciences,
Department of Electrical and Computer Engineering &
Center for Bioinformatics and Computational Biology
http://bioinformatics.udel.edu/Seminars/Current

MONDAY, September 26, 2016
3:30pm
DBI Room 102

Using Machine Learning Techniques to Analyze Electronic Health Records for Patients Risk Stratification

Claudine Jurkovitz

Senior Physician Scientist, Value Institute, Christiana Care Health System

Moumita Battacharya

PhD Student, Computer & Information Sciences

ABSTRACT: Chronic kidney disease (CKD) is prevalent in the United States. CKD patients have multiple comorbid conditions and have a high rate of hospitalizations. We seek to investigate the effectiveness of the health care delivery system in providing care for CKD patients in Delaware. To do so, we linked nonintegrated clinical data from multiple sources to create a longitudinal description of care. Based on these data, we are developing models for predicting hospital admissions.

Our first step is to identify patterns of association among co-occurring conditions. We will present a data-driven approach, applying a machine learning method, namely topic modeling, to Electronic Health Records (EHRs), aiming to identify patterns of associated conditions. We use the Latent Dirichlet Allocation (LDA), a method based on the idea that documents can be modeled as a mixture of latent topics, where each topic is a distribution over words. In our study, we adapt the LDA model to discover latent topics in patients' EHRs. The list of diagnosed conditions for each patient is viewed as a document, and each diagnosed condition is treated as a word. The latent topics obtained using LDA can be characterized by a distribution over diagnosed conditions, wherein each condition is assigned a probability to be associated with each topic. Our next step is to identify important predictors of the different stages of chronic kidney disease using supervised machine learning algorithms.