Biomedical Informatics Grand Rounds
Wednesday, Sept. 1, 2021 3:00 pm – 4:00 pm

Three recent vignettes in clinical informatics: time, language, and scale

Janos G. Hajagos, PhD
Research Assistant Professor and Chief of Data Analytics
Department of Biomedical Informatics, Stony Brook University
Stony Brook, NY

Remote Access
Join Zoom Meeting https://stonybrook.zoom.us/j/95617197636?pwd=KytzZ2pVRG9SZGpKZUtpNXJISjNjZz09
Meeting ID: 956 1719 7636 Passcode: 924293

Bio: Dr. Janos G. Hajagos is Chief of Data Analytics and Research Assistant Professor in the Department of Biomedical Informatics at Stony Brook University. He is the lead data scientist across several quality initiatives at Stony Brook Medicine and for the Suffolk Care Collaborative's DSRIP award. Before his current position, he was the Associate Director of Data Computation in the Division of Applied Informatics.

Dr. Hajagos received his Ph.D. in Ecology and Evolutionary Biology from Stony Brook University, NY (2005) and a B.A. in Biology from Bard College, NY (1999). He has published widely from risk analysis to applications of the semantic web to healthcare. His areas of research in biomedical informatics can be divided into two areas. The first is utilizing machine learning approaches to quantify patterns of health care utilization. This includes the application of graph databases and algorithms to health care data. The second is the application of semantic web technologies to bridge disparate sources of information. Dr. Hajagos has been developing tools to adapt and utilize the UMLS to find novel links between basic science researchers and clinicians.

Abstract: We will look at three recent informatics projects that span across different methodological approaches. The first project we will look at is how sequence-based models can be used to predict time to acute and chronic conditions diagnosis in hospitalized patients. Second, we will look at how radiology reports can be parsed and used in a task for predicting the diagnosis of pneumonia. Third, we will examine how data pipelines for mapping data to the OHDSI CDM are being scaled to run in the Apache SPARK environment.

Educational Objects: Upon completion, participants should be able to:
● Understand how sequence-based models can be used to predict acute conditions such as AKI with EHR (Electronic Health Records) and what advantage do they have over models that use traditional engineered features.
● Learn how a radiology report structure can be parsed and how the structure can be used to build prediction models using standard ML tools.
● The Apache SPARK environment allows data pipelines to be scaled across multiple machines. Learn how this can be exploited to map EHR data to the OHDSI CDM (Common Data Model) at scale.