Invited Talk: Analyzing Tandem Mass Spectra: A Graphical Models Perspective

John T. Halloran HALLOJ3@UW.EDU
University of California, Davis
Davis, California (USA)

Abstract

In the past two decades, the field of proteomics has seen explosive growth, largely due to the development of tandem mass spectrometry (MS/MS). With a complex biological sample as input, a typical MS/MS experiment quickly produces a large (often numbering in the hundreds-of-thousands) collection of spectra representative of the proteins present in the original complex sample. A majority of widely used methods to search and identify MS/MS spectra use scoring functions which rely on static, hand-selected parameters rather than affording the ability to learn parameters and adapt to the widely varying characteristics of MS/MS data. In this talk, we discuss recent work utilizing dynamic Bayesian networks (DBNs) to identify MS/MS spectra. In particular, we discuss a recently proposed DBN for Rapid Identification of Peptides (DRIP) which, in contrast to popular scoring functions, allows efficient generative and discriminative learning of parameters to achieve state-of-the-art spectrum-identification accuracy. Furthermore, facilitated by DRIP's generative nature, we present current innovations leveraging DBNs to significantly enhance many other aspects of MS/MS analysis, such as improving downstream discriminative classification via detailed feature extraction and speeding up identification runtime using trellises and approximate inference.

Biography

John T. Halloran is a postdoctoral researcher at the University of California, Davis. He completed his PhD in electrical engineering at the University of Washington, Seattle, in 2016. At the University of Washington, he worked in the MELODI (MachineE Learning for Optimization and Data Interpretation) and Noble Labs under the joint supervision of Jeff Bilmes and William Noble. He obtained his MS in electrical engineering in 2010 at the University of Hawaii, Manoa, and received a BS in electrical engineering and a BS in math from Seattle University in 2008. His current research interests lie in utilizing machine learning methods to analyze proteomics data, with particular focus on the development of graphical models which may be effectively trained (either generatively or discriminatively). His previous work has involved developing graphical models for cancer genomics, automatic speech recognition, and wireless communications. He is also broadly interested in machine learning applications and problems in computational biology.