SEMINAR NOTICE

Preston M. Green Department of Electrical and Systems Engineering

Machine Learning Morphisms: A Framework for Design and Optimization of Machine Learning Workflows

PhD Preliminary Research Examination

Eric Cawi
PhD Candidate

Preston M. Green Department of Electrical and Systems Engineering
Washington University in St. Louis

Abstract: At its heart, machine learning represents a sequence of transformations acting on data to extract relevant information or solve problems of interest. However, there is no unified language to express or analyze these transformations, and building applications with machine learning often relies on intuition or trial and error. In this presentation, we propose the concept of the Machine Learning Morphism (MLM) as a fundamental building block to express operations performed in machine learning such as data preprocessing, feature extraction, and model training. Inspired by statistical learning, MLMs are morphisms whose parameters are minimized via a risk function. We explore operations such as equality, composition, and decomposition of MLMs, which are used to build a machine learning workflow from data preprocessing to final task completion. The advantage of this framework lies in the ability to easily build, organize, and compare multiple workflows, and allows joint optimization of parameters across multiple steps in an application.

We apply the MLM to design a machine learning workflow that predicts 30 Day Hospital Readmissions using real data from Barnes Jewish Hospital. In this workflow, we first establish the Mapper Algorithm from Topological Data Analysis as an MLM, and then use it to train multiple models on subsets of the training data. These models then work as an ensemble to improve the prediction of high-risk patients compared to the current tools used at the hospital and out of the box machine learning tools. Using the MLM framework, we can describe this workflow with a single equation, and train the Mapper parameters jointly with the model parameters to optimize for ROC AUC on cross validation hold-out sets.

DATE: Tuesday, January 21, 2020
TIME: 12:00 p.m.
PLACE: Green Hall, Room 0120

Dissertation advisor:
Dr. Arye Nehorai

This seminar is in partial fulfillment
of the Doctor of Philosophy degree

Washington University in St. Louis
SCHOOL OF ENGINEERING & APPLIED SCIENCE