CoCo Seminar Series
Spring 2017

Machine Learning on the Structured Data: Convex Optimization for Group Feature Selection of Networked Data

Dr. Daehan Won
Assistant Professor, Department of Systems Science and Industrial Engineering
Binghamton University

Wednesday March 29th, 2017 8:30-9:30am
Engineering Building H-9 (Knoll-MacDonald Commons / Watson Commons)

In the machine learning, feature selection is one of the key concepts, and effective selection facilitates the interpretation of data and reduces computation costs as well as improving predictive performance. In data with structured features, there are relationships among features, e.g., features being connectivities between two nodes/sensors, and feature selection on the structured data has received a considerable attention in the past few years. Structured feature selection may reveal the complex structure of the data while providing compatible or better predictive power than state of the art algorithms. In this talk, I present a group feature selection problem on networked data, in which features are a collection of edge weights and feature groups are a collection of nodes in the network. The node selection problem entails an L-0 regularization, which is NP-hard because of the non-convex property. Most group feature selection and feature selection methods in the existing studies often avoid the L0 norm by replacing it with the convex L1 or L2 norm due to the combinatorics nature of L0 norm. To overcome the computational challenge of the `0 norm regularization in group feature selection, we propose a mathematical model with L0 norm based on the support vector machines (SVM) formulation and develop a heuristic-based computational technique to solve. This technique is based on a convex relaxation to reformulate the model as a semi-infinite program (SIP) and a new iterative algorithm that solves a sequence of multiple kernel learning sub-problems. Experimental results on synthetic and real human brain network datasets show that the proposed technique outperforms state-of-the-art group feature selection and feature selection methods as well as other baseline classification models without regularization. In addition, I propose exact solution approaches to solve the original problem as showing current researches on this topic.

Dr. Daehan Won is an Assistant Professor in the Department of Systems Science and Industrial Engineering at Binghamton University. His research interest is in large scale data analysis via mathematical programming and algorithms.

For more information, contact Changqing Cheng (ccheng@binghamton.edu). http://coco.binghamton.edu/