UNIVERSITY OF CALIFORNIA MERCED | School of Natural Sciences

APPLIED MATHEMATICS SEMINAR 291

Marginalized Graph Kernels for Learning on Molecules: Theory, Software, Stack, and Applications

Date: **10/25/19**

Time: **3:00 PM**

Location: **COB1 265**

Yu-Hang Tang

Lawrence Berkeley National Lab

For more information, contact : **Changho Kim**
ckim103@ucmerced.edu

Abstract

In this talk, I will discuss in three parts how kernel-based learning methods can be applied to graph-based datasets for building predictive models of molecular and atomistic properties that are accurate and interpretable using scarce training data.

In the first part, I will present an active learning protocol using Gaussian process regression, which prompts the need for similarity measures between either entire molecules or individual atomistic neighborhoods. This leads to our recent work on a kernel that operates on graphical representations of molecules. The graph kernel is intuitive and allows the application of the kernel trick to molecules of arbitrary size and topology.

In the second part, I will introduce GraphDot, a Python package that implements the marginalized graph kernel for and beyond molecules on general-purpose GPUs. GraphDot delivers thousands of times of speedups against existing CPU-only packages. This is achieved by taking advantage of a generalized Kronecker product structure in the linear algebra form of the graph kernel to fully exploit the GPU's tremendous floating point capability. A two-level sparse format and an adaptive primitive switching mechanism are used to ensure equal efficiency of the algorithm for both dense and sparse graphs.

In the last part, I will showcase specific applications using the graph kernel framework and the GraphDot package to rapidly construct Gaussian process regression models within minutes. With an active learning procedure, we can build GPR models to achieve an accuracy level of less than 1 kcal/mol for predicting atomization energies without using explicit energy decomposition/localization. More examples involving the direct prediction of experimental observables will also be given.