



Thesis Proposal

GHC 6501 | Monday, May 13, 2019 | 9:30 am

Learning with Kernels at Scale

Wei-Cheng Chang

Abstract

Over the past two decades, kernel methods have been one of the most active areas in machine learning, with a broad range of successful applications. For instance, kernels for computing similarities between structured objects in bioinformatics; covariance kernels in the Gaussian Processes to control inductive bias in time series modeling. Kernels also define mean feature map to measure the distance between two probability distributions. Despite the ubiquity, kernel methods often suffer from two fundamental limitations: the tractability of large-scale problems, and the difficulties in kernel selection/learning for complex downstream tasks.

The first part addresses the tractability challenge in kernel approximation, especially in the use of random Fourier features (RFF) to approximate shift-invariant kernels. We proposed a method for achieving a similar approximation accuracy with much smaller RFF sample, by learning the non-uniform weights for sampled RFFs from the data on hand. In addition, we aim to study advanced RFF-variants to improve the scalability of kernel contextual bandits problems

The 2nd part of the thesis focus on kernel selection/learning for the problem on hand. Choosing a suitable kernel is essential for getting good practical performance; however, how to find such a kernel for a given problem is often not obvious. we propose a novel framework, namely MMD-GAN, that optimizes kernel maximum mean discrepancy (MMD) parametrized by deep neural networks as a sensible loss to train deep generative models. On the other hand, we study kernel selections for change-point detection problem. We propose to optimize a lower bound of test power via surrogate generative models to sample sufficient anomalies.

The 3rd part of the thesis focuses on both the theoretic study and empirical examination of the relations between MMD and other popular distance metrics such as optimal transport and Sinkhorn divergence. Especially, we propose to apply MMD and Sinkhorn divergence to two downstream tasks, i.e., multi-label classification and multimodal word distribution. We hope our investigation on this part to shield new insights in this direction.

<https://drive.google.com/file/d/1vB4uW3MVBuEOgYMGMXMZPWTdBFM03Ki/view>



COMMITTEE:

Yiming Yang, (Chair)

Barnabas Poczos

Jeff Schneider

**Sanjiv Kumar
(Google Research)**