Thesis Defense

Zoom| Monday, August 3| 10:00 am

Learning with Kernels at Scale and Applications in Generative Modeling

Wei-Cheng Chang

COMMITTEE:

Yiming Yang, (chair)

Barnabas Poczos

Jeff Schneider

Sanjiv Kumar
(Google Research NYC)

Abstract

Kernel methods are versatile in machine learning and statistics. For instance, Kernel two-sample test induces Maximum Mean Discrepancy (MMD) to compare two distributions and serves as a distance metric for learning implicit generative models (IGMs). Kernel goodness-of-fit test, as another example, induces Kernel Stein Discrepancy (KSD) to measure model-data discrepancy and connects to a variational inference procedure for explicit generative models (EGMs). Other extensions include time series modeling, graph-based learning, and more. Despite their ubiquity, kernel methods often suffer from two fundamental limitations: the difficulty in kernel selection for complex downstream tasks and the tractability of large-scale problems. This thesis addresses both challenges in several complementary aspects.

In part I, we tackle the issue of kernel selection in learning implicit generative models (IGMs) with kernel MMD. Conventional methods using a fixed kernel MMD have limited success on high-dimensional complex distributions.

We propose to optimize MMD with neural-parametrized kernels, which is more expressive and improves the state-of-the-art results on high-dimensional distributions (Chapter 2). We also formulate kernel selection problems as learning kernel spectral distributions, and enrich the spectral distributions by learning IGMs to draw samples from it (Chapter 3).

In part II, we aim at learning suitable kernels for Stein variational inference descent (SVGD) in explicit generative modeling (EGMs). Although SVGD with fixed kernels shows encouraging performance in low-dimensional (within hundreds) Bayesian inference tasks, its success with high-dimensional problems such as image generation is limited. We propose the noise-conditional kernel SVGD (NCK-SVGD) for adaptive kernel learning, which is the first SVGD variant successfully scaled up for distributions with the dimension of several thousands, and performs competitively as state-of-the-art IGMs. With a novel entropy regularizer, NCK-SVGD enjoys flexible control between sample diversity and quality (Chapter 4).

In part III, we address the kernel tractability challenge with variants of random Fourier features (RFF). We propose to learn non-uniformly weighted RFF, which performs as good as the uniformly-weighted RFF while demanding less memory (Chapter 5). For the kernel contextual bandit problem, we reduce the computational cost of the kernel UCB algorithm by using RFF to approximate the predictive mean and confidence estimate (Chapter 6).

In part IV, We extents kernel learning for time series modeling and graph-based learning. For change-point detection over time series, we optimize the kernel two-sample test via auxiliary generative models, acting as surrogate samplers of unknown anomaly distributions (Chapter 7). For graph-based transfer learning, we construct the graph Laplacian by kernel diffusion and leverage label propagation to transfer knowledge from source to target domains (Chapter 8).

https://drive.google.com/file/d/1h9qlClGriUL6BbOUUKzWTATXX-e1gp2f/view