On the Curse of Memory in Recurrent Neural Networks: Approximation and Optimization Analysis

Speaker: Zhong Li (Peking University)
Time: 2021-10-18, 16:10 to 17:00
Location: Rm 1801, Guanghua East Tower
Advisor: Weinan E (Beijing Institute of Big Data Research)

Abstract: Recurrent neural networks (RNNs) are among the most frequently employed methods to build machine learning models on temporal data. Despite its ubiquitous applications, many fundamental theoretical questions remain to be answered. We study the approximation properties and optimization dynamics of RNNs when applied to learn input-output relationships in temporal data. We consider the simple but representative setting of using continuous-time linear RNNs to learn from data generated by linear relationships. Mathematically, the latter can be understood as a sequence of linear functionals. We prove a universal approximation theorem of such linear functionals and characterize the approximation rate. Moreover, a fine-grained dynamical analysis of training linear RNNs by gradient methods is performed. A unifying theme uncovered is the non-trivial effect of memory, a notion that can be made precise in our framework, on both approximation and optimization. When there is long-term memory in the target, it takes a large number of neurons to approximate it. Moreover, the training process will suffer from severe slow downs. In particular, both of these effects become exponentially more pronounced with increasing memory—a phenomenon we call the "curse of memory". These analyses represent a basic step towards a concrete mathematical understanding of new phenomenons that may arise in learning temporal relationships using recurrent architectures.