FUDAN UNIVERSITY

Applied Math Ph.D. Seminar

Initialization is Critical to Whether Transformers Fit Composite Functions by Inference or Memorizing

Speaker: Zhongwang Zhang (Shanghai Jiao Tong University)

Time: 2024-10-10, 16:10 to 17:00

Location: Rm 1801, Guanghua East Tower

Advisor: Zhi-Qin John Xu (Shanghai Jiao Tong University)

Abstract: Transformers have shown impressive capabilities across various tasks, but their performance on compositional problems remains a topic of debate. In this work, we investigate the mechanisms of how transformers behave on unseen compositional tasks. We discover that the parameter initialization scale plays a critical role in determining whether the model learns inferential solutions, which capture the underlying compositional primitives, or symmetric solutions, which simply memorize mappings without understanding the compositional structure. By analyzing the information flow and vector representations within the model, we reveal the distinct mechanisms underlying these solution types. We further find that inferential solutions exhibit low complexity bias, which we hypothesize is a key factor enabling them to learn individual mappings for single anchors. Our findings provide valuable insights into the role of initialization scale in shaping the type of solution learned by transformers and their ability to learn and generalize compositional tasks.