



面向多租户分布式机器学习的聚合传输协议

Wenfei Wu

Tsinghua University

Time: 11:00-12:00 (Time in Beijing)
15:00-16:00 (Time in Auckland)

July 16, 2021 (Friday)

Venue: B1-501, Main Building

Abstract: 随着机器学习数据集和模型的增大，机器学习的训练过程逐步被分布式部署到多服务器上，其中多worker向参数服务器PS交换梯度、更新模型的计算方式是一种典型的体系结构。但是，在这种体系结构下，PS容易成为通信瓶颈。我们设计了聚合传输协议ATP来解决这一瓶颈，同时支持在数据中心中的多租户多机柜部署。ATP利用最近的可编程交换机技术，将参数聚合的过程卸载到交换机上，从而减小了PS的网络流量和计算量。ATP协议包括交换机上的网内聚合计算服务、终端服务器的可靠传输、和高吞吐网卡的加速技术。我们将ATP对接PyTorch并在AlexNet、VGG等常用模型上进行测试，证明ATP能够有效的加速机器学习的效率。

Speaker Bio: 吴文斐，清华大学任助理教授。2015年博士毕业于美国威斯康星大学麦迪逊分校，后在惠普实验室任博士后研究院。2017年加入清华大学工作至今。SIGCOMM、NSDI、INFOCOM等网络顶级会议上发表论文30余篇，拥有美国专利3项。获SoCC13最佳学生论文、IPCCC最佳论文提名。

Chengdu Algorithms and Logic Seminar is a series of online seminars organized by School of Computer Science and Engineering, University of Electronic Science and Technology of China, and School of Computer Science, University of Auckland that aims to promote collaborations in a broad range of topics in algorithms and logic.

Organizers: Bakhadyr Khossainov, Jiamou Liu, and Mingyu Xiao

Email: myxiao@gmail.com (Mingyu Xiao)