Search

CENTER FOR COMPUTATION & TECHNOLOGY
Interdisciplinary | Innovative | Inventive

LOUISIANA STATE UNIVERSITY

Home | About ▼ | Focus Areas ▼ | Research ▼ | Divisions ▼ | Programs ▼ | News ▼ | Events ▼ | Computing Resources ▼ | Jobs | Contact Us | Log In | LSU ▼

Events

Current Events
Lectures▼
Events Archive▼

Other - Colloquium on Artificial Intelligence Research and Optimization

Rethinking the Parallelization of Extreme-scale Deep Learning

Abhinav Bhatele, University of Maryland, College Park

Associate Professor

Hybrid- Digital Media Center/Zoom Theatre
October 05, 2022 - 03:00 pm

Abstract:

ZOOM INFO:
Webinar ID: 927 6041 9250
Passcode: 116287

The rapid increase in memory capacity and computational power of modern architectures, especially accelerators, in large data centers and supercomputers, has led to a frenzy in training extremely large deep neural networks. However, efficient use of large parallel resources for extreme-scale deep learning requires scalable algorithms coupled with high-performing implementations on such machines. In this talk, I will present AxoNN, a parallel deep learning framework that exploits asynchrony and message-driven execution to optimize work scheduling and communication, which are often critical bottlenecks in achieving high performance. I will also discuss different approaches for memory savings such as using CPU memory as a scratch pad, and magnitude-based parameter pruning. Integrating these approaches with AxoNN enables us to train large models using fewer GPUs, and also helps reduce the volume of communication sent over the network.

Speaker's Bio:

Abhinav Bhatele is an assistant professor in the department of computer science, and director of the Parallel Software and Systems Group at the University of Maryland, College Park. His research interests are broadly in systems and networks, with a focus on parallel computing and large-scale data analytics. He has published research in parallel programming models and runtimes, network design and simulation, applications of machine learning to parallel systems, parallel deep learning, and on analyzing/visualizing, modeling and optimizing the performance of parallel software and systems. Abhinav has received best paper awards at Euro-Par 2009, IPDPS 2013 and IPDPS 2016. Abhinav was selected as a recipient of the IEEE TCSC Young Achievers in Scalable Computing award in 2014, the LLNL Early and Mid-Career Recognition award in 2018, and the NSF CAREER award in 2021. Abhinav received a B.Tech. degree in Computer Science and Engineering from I.I.T. Kanpur, India in May 2005, and M.S. and Ph.D. degrees in Computer Science from the University of Illinois at Urbana-Champaign in 2007 and 2010 respectively. He was a post-doc and later computer scientist in the Center for Applied Scientific Computing at Lawrence Livermore National Laboratory from 2011-2019.