“Algorithmic Innovations for Accelerating Implementations of Artificial Intelligence”

Abstract: Artificial Intelligence and Machine Learning (AI/ML) are playing a central role in pushing the frontiers of technology and transforming our society. However, with rapid innovation in these techniques, the complexity and scale of the problem being solved is increasing dramatically. Another trend that we are witnessing is the dramatic increase in the heterogeneity of computing systems for AI/ML processing. The end of Moore’s law coupled with the proliferation of AI/ML workloads in almost every aspect of the society has forced hardware developers to come up with novel heterogeneous hardware solutions similar to System-on-Chip (SOC) paradigms. For a data scientist, learning the intricacies of mapping applications to various hardware devices is not only challenging but also counterproductive as the time can be better spent in innovations in the applications itself.

In this talk, I will present my research towards developing frameworks that enable high performant implementations of various AI algorithms on a variety of heterogeneous platforms. Using Deep Reinforcement Learning as an example, I will illustrate how a deep understanding of the computational characteristics of the algorithms, combined with an ability to develop analytical models of underlying platforms can enable us to develop better performing algorithms on a variety of platforms, while ensuring that the hardware is well utilized. I will then discuss my future research directions and identify research interests that may be complementary to the framework and can add to its efficacy.

Bio: Sanmukh Rao Kuppannagari is an Assistant Professor at the Computer and Data Sciences Department at Case Western Reserve University. Prior to this, he was a Senior Research Associate at the University of Southern California (USC). He completed his PhD in 2018 in Computer Engineering from USC. He holds a Bachelor of Technology degree from the Indian Institute of Technology, Guwahati in Computer Science and Engineering. His research interests include reconfigurable computing, combinatorial optimization, approximation algorithms, and AI/ML acceleration. He is a Co-PI on four funded National Science Foundation awards and one Army Research Office Award. He has published more than 50 papers with a couple of them receiving “Best Student Paper” awards.