CHALLENGE #3
CAT-LEA-01

Distributed Deep Learning with Edge Computing

Meet the expectations of this US Node through the technology challenge described below

APPLY NOW

GOALS

The aim of this project is to develop novel schemes to address one or more challenges described above to enable distributed AI/ML in edge computing clouds. The Explorers also collaborate with faculty and graduate students to work on a NSF-funded project on mobile edge computing.

DETAILS

Edge servers deployed in radio access networks as well as user mobile terminals form an edge cloud which can process data and perform computation in close proximity to end users with low latency. Artificial Intelligence (AI)/Machine Learning (ML) become more and more important tools to process and analyze big data, make predictions, and determine the mathematical optimization in real time. Machine learning algorithms are used in a wide variety of real-time edge applications, such as augmented reality/virtual reality, robot control, and autonomous driving. In order to effectively support distributed AI/ML applications on mobile terminals, e.g. smartphone, smart car, mobile robot, UAV as well as edge servers, several challenges should be addressed, including (1) how should mobile terminals and edge servers cooperate to preform distributed AI/ML applications with low latency and energy efficiency? (2) How to train a distributed AI/ML model across an edge network? (3) How to reduce communication overhead for distributed AI/ML applications in distributed/federated training and inference process? (4) How to update AI/ML model and distribute intermediate output among mobile terminals and edge servers based on the changing task and environment? (5) Where and how to place the distributed AI/ML application services in an edge computing cloud? (6) How to schedule the AI/ML tasks and allocate resources?

F6S · zabala innovation consulting · AUSTRALO