ITW-22 Themed Session: Communication-efficient Gradient Compression and Coding in Distributed Learning

Call for Student Research Presentations

The digitalization of physical devices, system infrastructure, and data services, as embodied by the IoT paradigm, has enabled the collection of large-scale databases which can be used for a trove of machine learning (ML) tasks, ranging from autonomous driving to health-care services and smart energy management. In this paradigm, data centralization is no longer a feasible option and thus distributed ML is being hailed as the next milestone in large-scale data computing. In this context, gradient compression and coding has emerged as a crucial bottleneck in the ability of training ever-larger deep neural network (DNN) models over an ever-large network of users. In the literature, authors often assume that remote users are connected to the parameter server (PS) through noiseless channel so that the concern is with respect to the network connectivity, client availability or model update scheduling. It is the organizers' opinion that this approach is somewhat insufficient in addressing the need for drastic reduction in payload required by modern-day distributed training scenarios. We believe that a renew information theoretical analysis could, instead, bring forth novel approaches. By viewing the gradient evaluated at each remote user as a information process with certain information content, it is possible to derive classical rate-distortion and channel coding approaches that can be applied to the distributed learning scenario. We believe that this approach has the potential of drastically improving the overall learning performance, as measured by the per-bit accuracy, that is the accuracy attainable after a certain number of bits have been exchanged over the network.

Potential topics:

Potential topics in this ITW themed session include, but are not limited to

- Communication-efficient gradient quantization, compression, and sparsification
- Statistical model of gradient evolution process and distribution-aware federated learning
- Distributed learning with stragglers or delayed model updates
- Convergence analysis with imperfect model update
- Multiple access techniques and coding for distributed learning

- Multi-terminal coding, such as over-the-air computation, for distributed learning
- Gradient attack, injection, and other security issues
- Privacy in distributed learning
- Network optimization for federated learning
- Distributed reinforcement, online, and sequential learning
- Gradient representation for communication and computation
- Application of distributed learning in communication, networking, and 3GPP AI study item

Presentation format : full 30 minutes video (offline) with a 5 minutes elevator pitch (live)

Submission Guideline: Please send a extended abstract (one page, two-column) to sclin2@ntu.edu.tw

Deadlines:

Extended abstract submission deadline : 2 Sep, 2022

Acceptance Notification: 30 Sep, 2022

Camera Ready videos: 21 Oct, 2022

Chairs:

Yu-Chih Huang, National Yang-Ming Chiao-Tung University, Taiwan
(jerryhuang@nctu.edu.tw)
Shih-Chun Lin, National Taiwan University, Taiwan
(sclin2@ntu.edu.tw)
Stefano Rini, National Yang-Ming Chiao-Tung University, Taiwan
(stefano@nctu.edu.tw)