Thesis Title:

Cooperative inference and learning for Internet-of-Things with limited resources

Abstract:

In the era of big data and Internet-of-Things (IoT), ubiquitous smart devices not only continuously sense the environment and generate large amount of data, but also work cooperatively as sensor networks to extract meaningful insight from sensed information. In order to better address the challenges arise from online inference and learning applications, such as system latency, limited battery power and communication bandwidth, decentralized adaptation approaches distribute computation and data storage resources to every device or nodes in the network, which allows information to be processed and fused through only local cooperation, and also improves robustness and scalability. Therefore, decentralized methods have become especially promising and popular for IoT applications compared with traditional centralized solutions or cloud-based architectures.

In this dissertation, we study decentralized inference and learning for the IoT from perspectives of both theoretical algorithms and engineering applications. To be specific, our contributions are as follows:

- Firstly, we consider a multitask estimation problem where nodes in a network are divided into several connected clusters, with each cluster performing a least-mean-squares estimation of a different random parameter vector. Inspired by the adapt-then-combine diffusion strategy, we propose a multitask diffusion strategy whose mean stability can be ensured whenever individual nodes are stable in the mean, regardless of the inter-cluster cooperation weights. In addition, the proposed strategy is able to achieve an asymptotically unbiased estimation, when the parameters have same mean. We also develop an inter-cluster cooperation weights selection scheme that allows each node in the network to locally optimize its inter-cluster cooperation weights to achieve a lower average steady-state network mean-square deviation (MSD).
- Second, we also consider the issues of preserving energy budget and bandwidth resources for a single-task sensor network. To this end, we proposed an event-based communication mechanism for diffusion least mean-squares estimation over networks, in which an intermediate estimate from a sensor is communicated to its neighbors only when a triggering criterion is satisfied. By applying this event-based strategy, the network can achieve similar steady-state network MSD as the adapt-then-combine diffusion strategy but at a significantly lower communication rate.

- Finally, we study an engineering application problem of decentralized deep learning for face identification. We adopt and implement the concept of edge-computing where every camera node is equipped with embedded computing platform to learn and extract features from video frames locally, followed by a local feature fusion and inference performed on an edge gateway device. Our proposed method and architecture alleviates the need of super powerful GPU-based server, and is able to reduce inference latency and the usage of communication bandwidth while being more scalable and robust compared to cloud-based approaches which is most commonly adopted.