Title: A Training Framework and Architecture Design for Distributed Deep Learning

Speaker: Mr Wang Wei

Date/Time: 20 January 2017, Friday, 09:00 AM to 10:30 AM

Venue: MR1, COM1-03-19

Supervisor: Dr Ooi Beng Chin, Professor, School of Computing

Abstract:

Deep learning has recently gained a lot of attention on account of its incredible success in many complex data-driven applications, such as image classification. However, deep learning is quite user-hostile and is thus difficult to apply. For example, it is difficult to implement a deep learning model from scratch, and it is tricky and slow to train a large model which may consume a lot of memory. In this talk, I will introduce our investigation and approaches to resolve related challenges.

First, we have done a comprehensive analysis of optimization techniques for deep learning systems. In particular, some techniques from database systems would be discussed, including the techniques of memory management and operation scheduling. Issues related to communication, consistency and fault-tolerance in distributed training will also be studied. Second, we have designed and developed a distributed deep learning system, named SINGA, which tackles the usability problem and realizes optimization techniques for distributed training discussed in the first part. SINGA provides a flexible system architecture for running different distributed training frameworks. Last, we have proposed deep learning based methods for effective multi-modal retrieval on top of SINGA, which outperform state-of-the-art approaches.