香港浸會大學
HONG KONG BAPTIST UNIVERSITY

DEPARTMENT OF
COMPUTER SCIENCE
計算機科學系

DEPARTMENT OF COMPUTER SCIENCE

SEMINAR

2025 SERIES

Modular Sparsification of DNNs to Improve Pruning Performance and Model Interpretability

DATE & TIME

13 JUN 2025 (FRI) 3:00 - 4:00 PM

VENUE

DLB111, 1/F, DAVID C. LAM BUILDING, SHAW CAMPUS

PROF. PING CHEN

Professor
Department of Engineering
University of Massachusetts Boston

ABSTRACT

Modern DNNs often include a huge number of parameters that are expensive for both computation and memory. Pruning can significantly reduce model complexity and lessen resource demands, and less complex models can also be easier to explain and interpret. In this paper, we propose a novel pruning algorithm, Cluster-Restricted Extreme Sparsity Pruning of Redundancy (CRESPR), to prune a neural network into modular units and achieve better pruning efficiency. With the Hessian matrix, we provide an analytic explanation of why modular structures in a sparse DNN can better maintain performance, especially at an extreme high pruning ratio. In CRESPR, each modular unit contains mostly internal connections, which clearly shows how subgroups of input features are processed through a DNN and eventually contribute to classification decisions. Such process-level revealing of internal working mechanisms undoubtedly leads to better interpretability of a black-box DNN model. Extensive experiments were conducted with multiple DNN architectures and datasets, and CRESPR achieves higher pruning performance than current state-of-the-art methods at high and extremely high pruning ratios. Additionally, we show how CRESPR improves model interpretability through a concrete example.

SPEAKER'S BIOGRAPHY

REGISTER NOW

Enquiries: 3411-2385 **Email:** comp@comp.hkbu.edu.hk **Website:** https://bit.ly/bucs-events