Department of Computer Science
香港城市大學
City University of Hong Kong

**COMPUTER SCIENCE COLLOQUIUM**

# Modular Sparsification of DNNs to Improve Pruning Performance and Model Interpretability


SPEAKER **Prof. Ping Chen**

Professor
Computer Engineering and the Director of Artificial Intelligence Lab at the University of Massachusetts Boston


DATE 21 Mar, 2025 (Fri)
TIME 4:00PM – 5:00PM
VENUE CS Seminar Room, Y6405, 6th Floor, Yellow Zone, Yeung Kin Man Academic Building, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong

## 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.


## BIOGRAPHY

Prof. Ping Chen is a Full Professor of Computer Engineering and the Director of Artificial Intelligence Lab at the University of Massachusetts Boston. His research areas include Natural Language Processing, Machine Learning, and Artificial Intelligence. He has published over 130 papers in major Natural Language Processing, Data Mining, Artificial Intelligence, and Computational Linguistics conferences and journals, such as ACL, NAACL, IJCAI, IEEE Intelligent Systems, Artificial Intelligence Journal, IEEE TKDE. Prof. Ping Chen received his BS degree on Information Science and Technology from Xi'an Jiao Tong University, MS degree on Computer Science from Chinese Academy of Sciences, and Ph.D degree on Information Technology at George Mason University.

**All are welcome!**

*In case of questions, please contact Prof. WANG Jianping at jianwang@cityu.edu.hk, or visit the CS Departmental Seminar Web at https://www.cs.cityu.edu.hk/events/cs-seminars/recent-cs-colloquiums.*