

國立中央大學數學系

專題演講

主 講 人：陳彥賓博士 (中研院統計所)

演講題目：**Cluster Images by Self-supervised Learning with a Fixed Embedding**

演講茶會：2022年04月21日(星期四) 3:30 p.m. ~ 4:00 p.m.

茶會地點：中央大學鴻經館M306

演講時間：2022年04月21日(星期四) 4:00 p.m. ~ 5:00 p.m.

演講地點：中央大學鴻經館M107

Abstract：

Deep neural network (DNN) models are being investigated for clustering unlabeled image data. Most existing DNN approaches train the model by inputting the training data, and preserve the data features in terms of embeddings. Those embeddings are certain dependent on their training data. This may raise the concerns such as generalizability, transferability, privacy, and security. Although self-supervised learning has made rapid progress, data clustering with a fixed embedding, which is low dependent on the training data, is rarely explored. In this talk, I will introduce our proposed Merge & Expand (ME) algorithm to cluster image data by using a fixed embedding and a DNN classification model. In brief, the proposed ME algorithm groups the embedded data into several small and tight clusters termed regions. Then, select the quite dissimilar regions as the seed regions along with the tentative labels. It merges similar seed regions, and expand the region size according to the outcomes by DNN's assistance. Finally, assigns cluster memberships. We further introduce into an idea of "clean" and "unclean" images in the clustering problem. ME reports the "clean" and "unclean" images where their geometric relations in the embedded space are compatible and incompatible with their cluster structure respectively. ME achieves a compatible performance on clustering unlabeled image data while comparing to other existing works.

敬 請 公 告 ， 歡 迎 參 加