

## CENTRE FOR SYSTEMS INFORMATICS ENGINEERING POSTDOC ASSOCIATES AND STUDENTS SEMINAR (CSIE PASS)

### Knowledge-informed Sparse Learning for Relevant Feature Selection and Optimal Quality Prediction

**Mr Yiren LIU**

Ph.D. Student, School of Data Science  
City University of Hong Kong



.....  
Date: 15 May 2023

Time: 3:00pm - 3:45pm

Zoom meeting ID: 994 8492 4425

Seminar link: <https://cityu.zoom.us/j/99484924425>

### Abstract

Industrial data are usually collinear, which can cause pure data-driven sparse learning to deselect physically relevant variables and select collinear surrogates. In this talk, we will introduce a novel two-step learning approach to retaining knowledge-informed variables (KIV) to build inferential models. The first step is an improved knowledge-informed Lasso (KILasso) algorithm by removing penalty on the KIVs to produce a series of candidate subsets that guarantee the retention of the KIVs. The candidate subsets are then used to run the KILasso or ridge regression again to select the best sets of variables and estimate the final model. Two new algorithms are proposed and applied to datasets from an industrial boiler process and the Dow Chemical challenge problem. It is demonstrated that some important physically relevant variables are deselected by pure data-driven sparse methods, but they are retained using the proposed knowledge-informed methods with superior prediction performance.

### Biography

Mr Yiren Liu is a third-year Ph.D. student with the School of Data Science, City University of Hong Kong. He received a Master degree in Financial Engineering from the University of Southern California, after a Bachelor degree in Software Engineering from Shandong University. His research interests include statistical machine learning, and sparsity statistical learning.