The Chemist's Interactions

Seminars @ the Chemistry Department

ROOM G11 h 14.30

Friday, 5th May 2023

Luigi Cavallo

Physical Science and Engineering Division
KAUST Catalysis Center (Saudi Arabia)

A Multivariate Linear Regression Workflow for Catalysts Design with Few-Entry Datasets

Catalysis is one of the technologies contributing to build a sustainable world by solving some of the great challenges humankind is facing. Considering the rate at which planet resources are consumed and climate is changing, there is urgency to act in the next decades before irreversible planetary damage will occur. Machine learning (ML) is hailed as a tool that could help faster development of better performing catalysts. This vision rests on the capability of ML tools to predict which catalyst structures outperform known ones. One critical point is the availability of large training datasets, as in most of the cases only few dozens of catalysts are available, and synthesis of new ones can require remarkable efforts. However, the accuracy of ML models trained on few-entry datasets to identify improved catalysts is unclear. In this work we introduce a series of confidence scores aimed to estimate the expected prediction capability of a new model. To achieve robust validation of the scores we applied the protocol to a variety of different cases. We thus assembled 512 experimental data corresponding to 29 individual reaction classes and experimental properties into 29 training datasets, and we trained and tested ML models to identify the top performing catalysts in each dataset. Due to the small size of these datasets, we rested on classic multivariate linear regression (MLR) as ML engine.

zoom

The event will be streamed on zoom.us for external participants!

For registrations: marco.ortenzi@unimi.it

UNIVERSITÀ DEGLI STUDI DI MILANO