Tata Institute of Fundamental Research
Survey No. 36/P, Gopanpally Village, Serilingampally, Ranga Reddy Dist., Hyderabad – 500 046

Seminar

Local Representation for Prediction of ^{13}C-NMR Chemical Shifts with Kernel Ridge Regression

Surajit Das

TIFR, Hyderabad

The integration of machine learning (ML) into nuclear magnetic resonance (NMR) spectroscopy has helped advance many aspects of structure elucidation in molecules and materials. To improve the accuracy of ML-based predictions, a primary strategy is to develop more precise representations, or descriptors, that effectively map molecular and atomic environments to local atomic properties. This method of encoding is known as the atoms-in-molecules scheme. In this study, a local descriptor is used to predict ^{13}C-NMR chemical shifts for small organic molecules in the QM9NMR dataset. This approach evaluates the distinction between discrete and continuous representations while implementing a nearest-neighbor inclusion concept. The aBoB-RBF(n) descriptors, with n = 3 and n = 4, yield significantly improved results compared to previously reported models. This highlights the importance of incorporating neighborhood information to accurately map chemical shift properties. Testing on the QM9 dataset shows that the aBoB RBF(4) model achieves a mean absolute error (MAE) of 1.69 ppm, outperforming previously reported methods.

With the Δ-ML approach, the MAE reduces to 1.16 ppm, approaching the 1 ppm error threshold required for the structure-elucidation benchmark. Further tests on external datasets, such as Drug12 and Drug40, confirm that the model is robust and performs well even for larger molecules not present in the training data. Finally, this model is implemented as a practical software module, *mlqm9nmr*, to predict ^{13}C-NMR chemical shifts.

Tuesday, Jun 30th 2026

11:30 Hrs (Tea / Coffee 11:15 Hrs)

Auditorium, TIFRH