Machine Learning and Statistical Analysis for Materials Science: Fingerprint Descriptors and Chemical Insights into Catalysts

Umesh V Waghmare
Theoretical Sciences Unit
Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR),
Jakkur P.O., Bangalore 560064 (India)
Email: waghmare@jncasr.ac.in

We first introduce computational high throughput screening for design and prediction of materials for a specific application, and argue the need for descriptors that can be readily calculated and used in quantitative description of the relevant material property. We demonstrate this through identification of structural and electronic descriptors of the catalytic activity of B and N doped graphene for the Oxygen Reduction Reaction (ORR). Using the resulting simple and predictive model, we identify the most optimal site of $C_{1-x}(B,N)_x$ for ORR [1]. In the paradigm of data-based approach to materials discovery, we have developed a semi-automated workflow or 'recipe' [2] that can help a material scientist to start from a raw dataset of materials' properties and descriptors, build predictive models, and draw insights into the governing mechanisms. We demonstrate our recipe, which employs machine learning tools and statistical analysis, through application to a case study leading to identification of descriptors relevant to catalysts for CO_2 electro-reduction, starting from a published database of 298 catalyst alloys. At the heart of our methodology, lies the Bootstrapped Projected Gradient Descent (BoPGD) algorithm that has significant advantages over commonly used Machine Learning (ML) and Statistical Analysis (SA) tools such as the regression coefficient shrinkage-based method (LASSO) or artificial neural networks. In addition to identifying the descriptors that parametrize the *d*-band model of catalysts for CO_2 reduction, we predict work function to be an essential and relevant descriptor, and propose a modification of the *d*-band model that includes the chemical effect of work function. Our scheme is general and particularly efficient in reducing a set of large number of descriptors to a minimal one, hence, we expect it to be a versatile tool in obtaining chemical insights into complex phenomena and development of predictive models for design of materials.

1. S Sinthika, U V Waghmare and R Thapa, Small 14, 1703609 (2018).

2. P Pankajakshan, S Sanyal, OE de Noord, I Bhattacharya, A Bhattacharyya, U V Waghmare Chemistry of Materials 29 (10), 4190-4201 (2017).

Work done in collaboration with *Praveen Pankajakshan, Suchismita Sanyal, Onno E. de Noord, Indranil Bhattacharya, Arnab Bhattacharyya, S Sinthika, R Thapa and Narendra Kumar*