New Techniques for Materials Space Exploration

Ganna Gryn'ova[1], Stiv Llenga[2]
[1]University of Birmingham, Edgbaston, B15 2TT
Birmingham, United Kingdom

g.grynova@bham.ac.uk

Rational exploration of the infinite space of hypothetical molecules and materials requires identifying the most important structural and compositional features for the target properties and applications. In this contribution, we discuss our recent efforts toward explainable machine learning for materials design, which include new quantum-inspired representations, fragment-based fingerprints, and dimensionality reduction techniques.

For example, we will introduce the matrix of orthogonalised atomic orbital coefficients (MAOC) representation [1], uniquely able to distinguish compounds with identical compositions and geometries but distinct charges and spin multiplicities. We will also discuss extensions of MAOC to extended and periodic systems. To greatly facilitate machine learning tasks on large datasets and involving extended materials, we will introduce the matrix of reference similarity (MRS) technique for reducing the dimensionality and representing the chemical space most compactly (Fig. 1). Finally, we demonstrate how new molecules and materials with optimal target properties (e.g., graphene quantum dots with low HOMO-LUMO gaps for organic electronics) can be rationally designed using the substructural filter representation (SFR), which identifies the building blocks with the highest feature importance.

References

[1] S. Llenga, G. Gryn'ova, *J. Chem. Phys.*, 158 (**2013**) 214116.

Figures

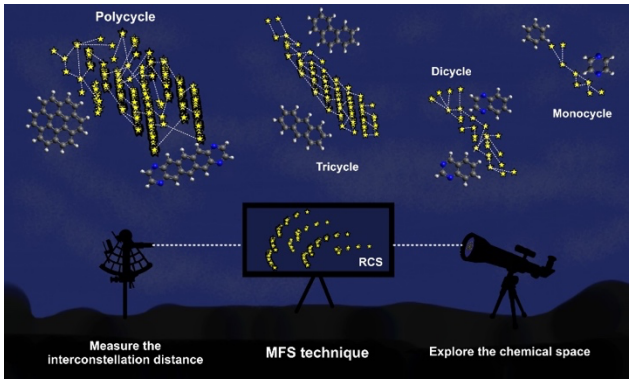

Figure 1. Mapping chemical space of N-heteropolycycles with the matrix of reference (fragment) similarity.