Recent Advances and Challenges in Machine Learning Representations of Molecules

Mairal Julien;[A]

A) Univ. Grenoble-Alpes, Inria.

Julien.mairal@inria.fr

The intersection of machine learning, cheminformatics, and computational biology has grown significantly, particularly through a branch of machine learning dedicated to the representation of molecular graphs. This convergence is driving innovations, or is expected to drive innovations, in applications such as drug discovery, materials science, and chemical property prediction.

In this talk, we will discuss two main classes of approaches that have emerged at the forefront: Graph Neural Networks (GNNs) and Transformers. Rather than listing the numerous variants, we will focus on their core principles and the current challenges these approaches face. These challenges often involve incorporating physical rules into the representations, such as 3D geometry, symmetries, and long-range potentials.

Additionally, we will explore the paradigms used to learn these representations. Self-supervised learning has become a key framework for designing foundation models in natural language processing and computer vision. By examining the ingredients from these fields, we aim to understand the difficulties preventing the development of a foundation model for chemistry.

KEYWORDS: **machine learning; graph neural networks**.