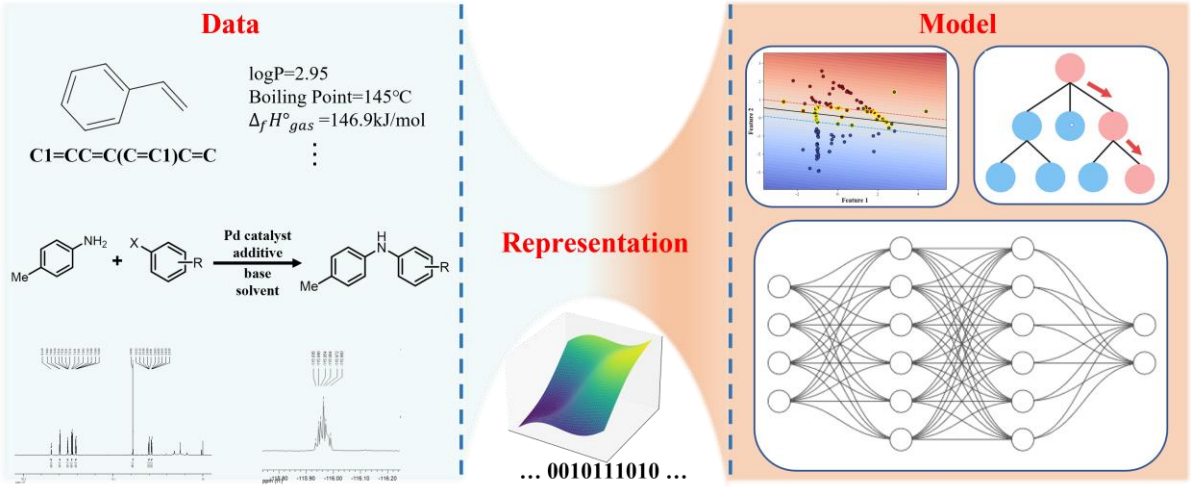
Survey on Recent Progress of AI for Chemistry: Methods, Applications, and Opportunities

Hu DING, Pengxiang HUA, Zhen HUANG

Frontiers of Computer Science, DOI: 10.1007/s11704-025-50127-3

Motivations & Ideas

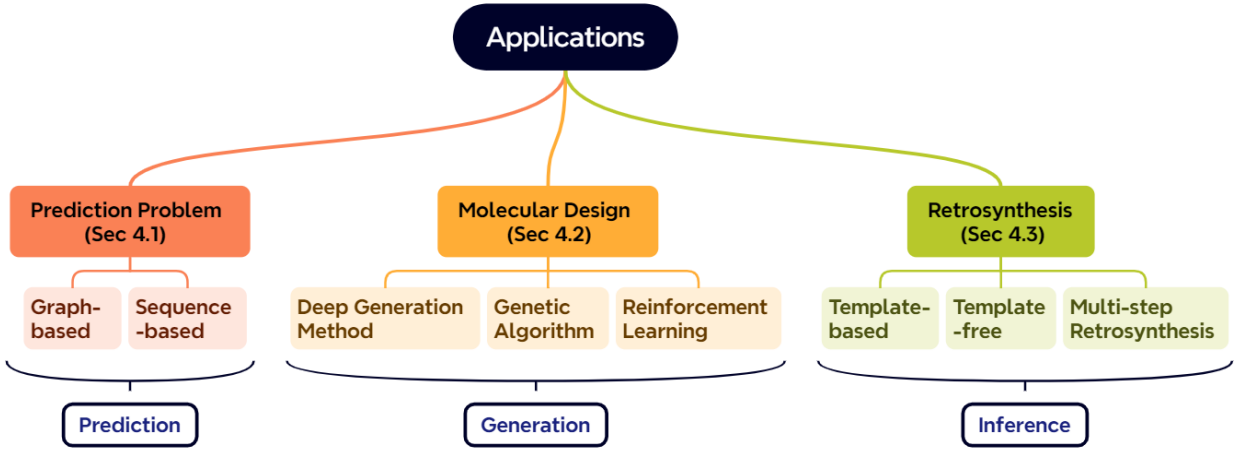
- Motivations:
  - The complexity of chemical knowledge presents a significant barrier for researchers from the computer science community.
  - Comprehensive overviews that cover the full pipeline of AI applications in chemistry from a computer science perspective remain limited.
- Ideas:
  - A Computation-Centric Framework: To provide a comprehensive review of *AI for Chemistry* from a computer science perspective, establishing a unified framework that covers Data, Representation, and Application.



Machine learning in chemistry involves three core components: **data**, **representation**, and **model**. Data provides raw chemical information; representation encodes it into machine-readable form; the model learns patterns to make predictions.

Main Contributions

- Contributions:
  - **A Unified Framework:** We review data sources and representation methods—from fingerprints to GNNs and Transformers—offering a structured roadmap for researchers.
  - **A Task-Based Taxonomy:** We propose a classification based on core ML tasks: prediction, generation, and inference, providing a computer science-oriented view of applications.
  - **Key Challenges Ahead:** We highlight four major challenges shaping future research: data scarcity, data bias, interpretability, and generative large models.



Three typical applications of AI for Chemistry.