Chair of Applied Computer Science X

Prof. Dr. Ruben Mayer

UNIVERSITÄT BAYREUTH

Title

Polymer prediction based on Graph Neural Networks

(Master Thesis)

Background

Polymers have emerged as a powerful class of materials for a wide range of applications because of their low-cost processing, chemical stability, tunable chemistries, and low densities. These attributes have led to vigorous, widespread, and sustained research, and to the development of new polymeric materials. The result is a constant flux of materials data. Over the past decade, the polymer informatics community has translated this data stream into machine-learned property predictors that efficiently screen libraries of candidate polymers for subsequent experimental inquiry.

Currently, most approaches for polymer screening rely on handcrafted features– extracted from the chemical structure of a polymer repeat unit–as input for property predictors. An alternative to handcrafting features is to "machine learn" them. One approach is to represent the material as raw text, such as a simplified molecular-input line-entry system (SMILES) or BigSMILES string, and then learn features with a neural network specifically designed for natural language processing. Another promising approach is to represent the material as a graph and then train a Graph Neural Network (GNN) to learn features. To date, GNNs have outperformed approaches based on handcrafted features.

Task definition

The goal of this thesis is to explore different ways of incorporating more detailed polymer graph models into GNN training. The crucial question is whether more detailed modeling can lead to higher accuracy at acceptable training costs.

The current state of the art in GNN-based polymer prediction is the Polymer Graph Neural Network (polyGNN) [1] architecture. While leading to competitive results, the modeling of polymer graphs in polyGNN is still very limited to simplified abstractions with the goal to keep the graphs small. Recent developments in scalable GNN training, e.g., via graph sampling and partitioning, make it possible to train GNNs on larger graphs. This paves the way to exploring more detailed polymer graph representations in GNN-based polymer prediction.

Literature: [1] Gurnani, R., Kuenneth, C., Toland, A., & Ramprasad, R. (2023). Polymer informatics at scale with multitask graph neural networks. *Chemistry of Materials*, *35*(4), 1560-1567.

Required Knowledge and Skills

- Basic knowledge in machine learning
- Programming skills (e.g., Python)

Supervisor

The thesis is carried out together under supervision Prof. Ruben Mayer (Data Systems) and Prof. Christopher Künneth (Materials Informatics) and involves interdisciplinary work at the intersection of computer science and materials science.

Feel free to ask questions anytime via Teams or e-mail (ruben.mayer@uni-bayreuth.de)

The Thesis can be written in English or German.