

Vorhersage von Bindemittleigenschaften: Eine Integration von Graph Neural Networks und Deskriptoren

Gaoyuan Zhang*, Christian Schmitz

Institut für Lacke und Oberflächenchemie, Hochschule Niederrhein, Adlerstraße 32, 47798 Krefeld

*E-Mail: gaoyuan.zhang@hs-niederrhein.de

Abstract

Motivation:

- Chemische Interpretation durch **KI**
- Um Bindemittleigenschaften zu modellieren, ist es notwendig sowohl lokale (z.B. funktionelle Gruppe) als auch globale (z.B. Molmasse) Bindemittleigenschaften für Lacke und Beschichtungen zu verstehen

GNN (Graph Neural Network):

- Bilden molekulare Strukturen effektiv als Graphen ab.
- Können geometrische Struktur und komplexe Wechselwirkungen innerhalb eines Moleküls darstellen.

Bindemittel-Deskriptoren:

- Quantitative oder kategoriale Messungen, die Gesamtheit der Bindemittelcharakteristiken beschreiben.

- Notwendig für eine übergeordnete Beschreibung der Polymereigenschaften wie Funktionalitäten und Mischungsverhältnisse.
- Bieten einen konsolidierten Überblick über die Eigenschaften von Polymeren.

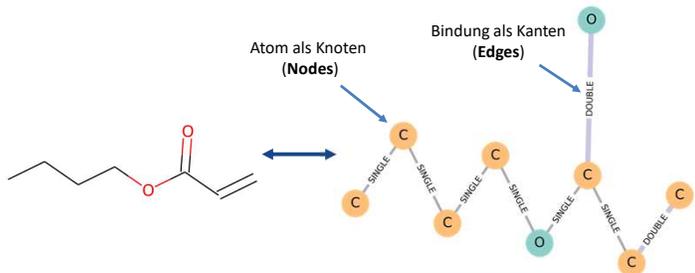
Neuerung:

- Einführung einer neuen multimodalen Methode: Verknüpfung von Molekülgraphen und Bindemittel-Deskriptoren.
- Löslichkeitsdatensatz** (~13.000 Daten) genutzt, um Graph- und globale Eigenschaften zu erstellen.
- Tg-Datensatz** von Homopolymeren mittels GNN
- Vergleich der Genauigkeit zwischen herkömmlichen Methoden und dem entwickelten **multimodalen** Modell.

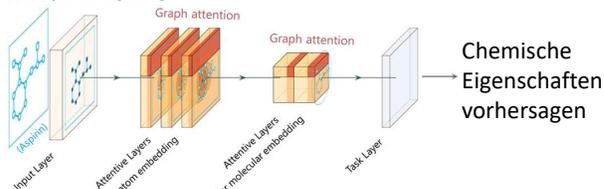
GNN Models

Moleküle als Graph

- Moleküle können als Graphen repräsentiert werden



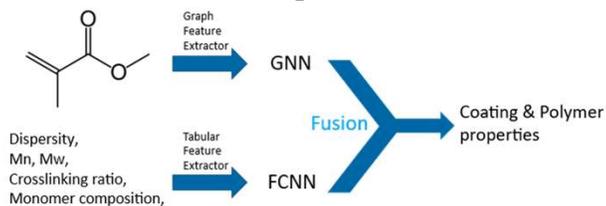
GNN für Chemie



- neuronales Netzwerk**, entwickelt für Graphen.
- Aktualisiert wiederholt Knoteninformationen, basierend auf den Informationen von benachbarten Knoten.
- Kann **lokale molekulare Eigenschaften** interpretieren und erkennt Zusammenhänge innerhalb der Graphen.

Multimodal-Fusion

Schematische Darstellung



- Weder die Moleküle noch die Deskriptoren allein genügen, um die Eigenschaften des Polymers zu beschreiben
- Die beiden NNs werden **fusioniert**, um Lack oder Polymer Eigenschaften vorherzusagen

Fusion Methode

Name	Beschreibung	Vorteil	Nachteil
Direkte Fusion	Direkte Training vom Fusion Model	Einfache Methode, Ohne Hyperparameter	Unausgeglichener Model-Trainingsprozess
Gradient Blending	Trainingsprozess ausgleichen durch Gradienten Mischung von mehreren Netzwerken	Ausgeglichener Modell-Trainingsprozess	Extra Training Aufwand, Hyperparameter-optimierung notwendig
OGM-GE*	Trainingsprozess regeln durch loss monitoring während Training	Trainingsprozess Ausgleicheung während Training	Applikation für Regression befindet sich nach wie vor in einem unreifen Stadium.

*Peng, Xiaokang, et al. *Balanced Multimodal Learning via On-the-Fly Gradient Modulation*. arXiv:2203.15332

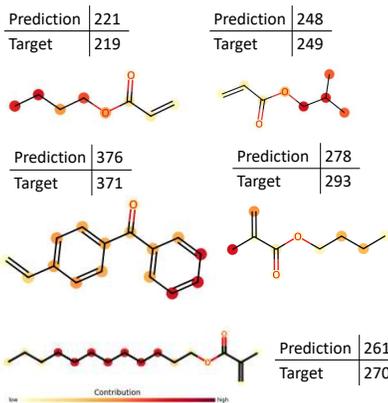
Ergebnis

- Applikationsbeispiel 1: **Multimodal**
- Vorhersagen Wasserlöslichkeit
- Datensatz: Wasserlöslichkeit (Log S) von ~13.000 kleinen Molekülen
- Fehler: Root Mean Squared Error (RMSE), Minimierung des Fehlers.

Model	Fehler (Log S)
QSAR Descriptor (FCNN) only	0.91
Molecule Graph (GNN) only	0.88
Direkte Fusion	>1*
Gradient Blending	0.85
OGM-GE	0.87

*Starker Overfit wegen unausgeglichener Trainingsprozess

- Applikationsbeispiel 2: **GNN**
- Monomereinfluss: Tg der Homopolymere



Fazit

- GNN erkennt Muster und Zusammenhänge in Molekülstrukturen
- Kombination von GNN und Deskriptoren verbessert Vorhersage
- Potenzial des multimodalen Ansatz zur Modellierung von Lackrohstoffen

Ausblick

- Erarbeitung eines Datensatzes zu chemischen Strukturen von UV-Lacken
- Weiterentwicklung und Optimierung des multimodalen Lernens und dessen Anwendung in der praktischen Lackformulierung für **KI-Anwendung**

Intelligente Infrastruktur für das Daten-Management in Automatisierten Chemischen Hochdurchsatz-Prozessen