

機械学習に基づく高熱伝導率ポリマーの設計事例

A case study on high thermal conductivity polymer design based on machine learning

データ科学研究系 ウ ステファン (Stephen Wu)

キーワード: 機械学習, マテリアルズ・インフォマティクス, ポリマー

1. はじめに

新しい材料を開発する従来の方法は、専門家の知識を使用して特定の物性要件を持つ少数の候補者を提案することです。このような方法は、膨大な化学空間（およそ 10^{60} 候補者 (Bohacek et al., 1996)）のために非効率的であることが知られている。人工知能技術の急速な発展のおかげで、機械学習に基づくより効率的な候補生成方法は、化学フロンティアの拡大ペースを加速すると期待されている。しかし、機械学習はどれほど強力であっても、単独で材料開発の挑戦を克服することはできません。計算分子設計法の開発の長い歴史にもかかわらず、合成までたどり着いたアルゴリズム的に設計された新規分子の例は限られている。これは、合成の難しさ、専門家の知識と学習された機械の知能の不一致、実用的なアプリケーションで厳しい要件などによるものであると推測しています。本研究では、Bayesian 分子設計という機械学習技術によって設計された高熱伝導率の新規ポリマーの発見を成功した。

2. 手法

ポリマーの熱伝導率 (λ) は実用的重要性のために多くの注目を集めている (Anderson, 1966)。しかし、設計のためにポリマーのメカニズムを完全には理解していない。本研究はこれまでに提案されているいくつかの計算設計法 (Zhong et al., 2001) と異なり、iqspr と呼ばれるベイジアン分子設計アルゴリズムに基づくデータ駆動アプローチを使用している (Ikebata et al., 2017)。データのソースは PolyInfo データベース (Otsuka et al., 2011) である。

R パッケージの iqspr は、ユーザーが指定した材料特性が与えられた探索空間を表す事後分布に対してサンプリングを実行します。ベイズの定理により、事後分布は尤度と事前分布との積として計算することができる。事前分布は、材料の物性に何の制約もなく化合物らしいものを表し、尤度は材料が与えられた場合に所望の材料物性を得る確率である。この問題設定の大きな課題は、信頼性の高い尤度モデルを訓練するためのデータが不十分であり、事後分布サンプリング・ステップ中に欠陥が生じることです。小さなデータの問題を克服するために、iqspr で元の目標の λ ではなく、 λ に関連する物性を選択する。図 1 は、PoLyInfo におけるいくつかの関連するポリマー物性のデータを示す。最終的に、高い λ の元のターゲットに代えて、高いガラス転移温度 (T_g) および熔融温度 (T_m) を目標とした。その後、iqspr から生成された提案候補は転移学習と専門知識を備えたポストスクリーニングを経た。我々の問題における転移学習は、 λ に相関し、十分に大きなデータセットを有する物性のモデルを構築して、これらの

モデルの一部は、 λ のモデル構築を助けるために利用される。そして、専門知識による産業応用の観点から設計の要素を追加する。

3. 結果

iqspr パッケージでは、 T_g が $> 250\text{ }^\circ\text{C}$ 、 T_m が $> 350\text{ }^\circ\text{C}$ になるようにターゲット領域を設定する。回帰モデルにはデフォルト値が使用され、 T_g 値と T_m 値が記録されたすべてのホモポリマーを含む訓練データが使用されます。ホモポリマーらしい分子を生成するために使用された従来のモデルでは、PoLyInfo データベースで利用可能な 14,424 ホモポリマーのすべてを使用した。その後、生成された分子はすべてポストスクリーニングを経て、三つの候補が専門家によって選択され、うまく合成された。図 2 は、我々の転移学習モデルからの予測値と比較した、3 つの候補の λ を示す。

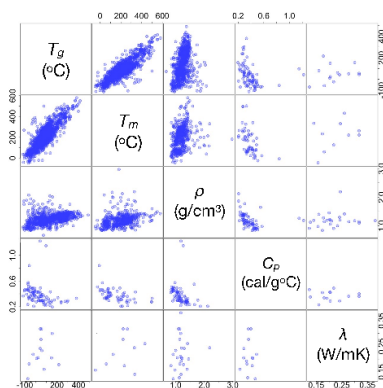


図 1. PoLyInfo 材料物性データの散布図
(ρ : 密度, C_p : 定圧比熱)

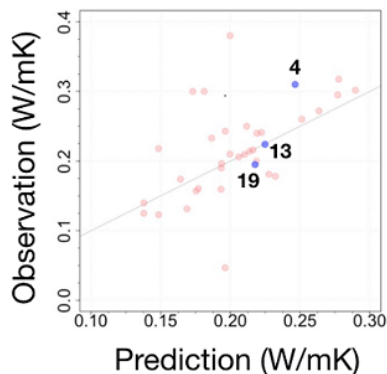


図 2. PoLyInfo の λ 観測値と転移学習モデルからの予測値との比較 (赤い点: 訓練データ, 青い点: 新しく合成したポリマー)

参 考 文 献

- Anderson, D.R. (1966). Thermal Conductivity of Polymers, *Chemical Reviews*, 66(6), 677-690.
- Bohacek, R.S., McMartin, C. and Guida, W.C. (1996). The art and practice of structure-based drug design: A molecular modeling perspective, *Medicinal Research Reviews*, 16(1), 3-50.
- Ikebata, H., Hongo, K., Isomura, T., Maezono, R. and Yoshida, R. (2017). Bayesian molecular design with a chemical language model, *Journal of computer-aided molecular design*, 31(4), 379-391.
- Otsuka, S., Kuwajima, I., Hosoya, J., Xu, Y. and Yamazaki, M. (2011). PolyInfo: Polymer database for Polymeric Materials Design, 2011 International Conference on Emerging Intelligent Data and Web Technologies, 22-29.
- Zhong, C., Yang, Q. and Wang, W. (2001). A group contribution model for the prediction of the thermal conductivity of polymer melts, *Industrial & Engineering Chemistry Research*, 40, 4151-4153.