

令和 5 年 6 月 20 日現在

機関番号：32660

研究種目：若手研究

研究期間：2021～2022

課題番号：21K14451

研究課題名（和文）畳込みグラフニューラルネットワークによる溶解度パラメータの多次元展開

研究課題名（英文）Development of high-dimensional Hansen solubility parameters by convolutional neural network

研究代表者

村上 裕哉（Murakami, Yuya）

東京理科大学・工学部工業化学科・助教

研究者番号：80880757

交付決定額（研究期間全体）：（直接経費） 3,600,000 円

研究成果の概要（和文）：本研究では、分子構造を入力値とした機械学習モデルの活用により、分子の特徴をフィンガープリント（FP）で表現し、基礎物性の推算を行った。大規模な量子計算データベースから事前学習を行う事で、分子の特徴を効率的に抽出しFPの作成に成功した。得られたFPにはHansen溶解度パラメータとの類似性が認められた。また、このFPを活用することで、沸点、臨界点、屈折率などの基礎物性の高精度な推算に成功した。本推算を実際の材料設計に活用するために、オルガノゲルとHansen溶解度パラメータとの間の相関関係の調査も行った。これらに相関がみられたことから、ゲル設計に作成したFPが活用可能であることが示唆された。

研究成果の学術的意義や社会的意義

本研究は、既存の量子計算に基づいたビッグデータからの事前学習を活用することで、分子構造からの物性推算を実現した。提案手法は、100データ程度の限られたデータ数からも既存のグループ寄与法を上回る精度で物性推算が可能であり、未知構造を有する分子の物性推算も可能であることが示された。これらの成果はから、本手法は物性推算分野における機械学習の有効的な利用法として、今後の発展が期待される。

研究成果の概要（英文）：In this research, molecular fingerprints were obtained using artificial intelligence, which takes a molecular structure as an input, and physical properties were accurately predicted from the obtained fingerprints. The characteristics of molecules were efficiently extracted from molecular structures by pre-training the model based on large database of quantum calculation. The obtained fingerprints have similarity to Hansen solubility parameters. Additionally, the fingerprints successfully predicted physical properties, including boiling point, critical points, and refractive index. To extend the usage of the obtained fingerprints, the correlation between organo-gel properties and Hansen solubility parameters was investigated. Since there was clear relationship between them, it is suggested that the obtained fingerprints can be used to tune the physical properties of organo-gels.

研究分野：化学工学

キーワード：溶解度パラメータ グラフニューラルネットワーク 機械学習 物性推算

## 1. 研究開始当初の背景

畳み込みグラフニューラルネットワーク (GCN) は、近年急速に開発が進んでいる機械学習モデルの一つであり、構造体を入力値とした推算などに広く用いられている。物性推算分野においては、分子構造を「グラフ」で表現することで入力値として利用可能であることが知られており、機械学習の高い表現力により高精度な推算を実現する手段として期待されている。<sup>[1]</sup>これにより従来のグループ寄与法などの経験則に基づいた推算手法の代替としての利用に向けて研究が進んでいる。

一方で、機械学習を活用するうえの一番の課題として、膨大なデータが必要な点が挙げられる。特に分子構造のような情報量が多い入力値を用いる場合、あらかじめ大量の教師データを用意する必要がある。さらに、溶解度などの複数分子が関与する物性の場合には、その組み合わせ数の指数関数的な増加に伴い、より豊富なデータ数が求められる。これらの課題を克服するためには、適切な事前学習による要求学習データ数の削減が有効である。特に近年は、量子計算によって得られた分子構造と分子物性の大規模データベースが利用可能となっており、これらの知見を活かすことで効率的な機械学習モデルの開発が可能になることが期待される。

本研究では、従来用いられてきた Hansen 溶解度パラメータに着想を得て、GCN の活用により分子を情報量に富んだパラメータ表現することを目指す。分子のパラメータ化には量子計算由来のビッグデータから学習を用い、注目する物性の推算時のみに実測データからの学習を行う事で、必要となる実測データ数の大幅な削減が見込まれる。

## 2. 研究の目的

本研究では、GCN の汎用性を拡張した Message Passing Neural Network (MPNN) を活用することで、分子の効率的なパラメータを行う。これにより、従来の手法では困難であった分子構造と物性の間に存在する黙示的な関連性を数値表現し、高精度かつ汎用性にすぐれた物性推算を可能とすることを目指す。この際に、(1)量子計算に基づいた既存のビッグデータからの事前学習、(2)Hansen 溶解度パラメータに着想を得た分子のベクトル表現を通じて、限られたデータ数からの物性推算を実現する。

また、得られたパラメータの実用的な活用を指向し、従来では分子構造から推算が困難であった物性の収集も行う。本研究では、特に有機溶媒と低分子量ゲル化剤から構成される超分子ゲルに着目し、作製材料の強度・光学的特性について Hansen 溶解度パラメータを用いた整理を行った。これらの結果を前述のパラメータと組み合わせることで、より高精度な推算が可能となることが見込まれる。

## 3. 研究の方法

MPNN は、python の PyTorch および PyGeometric ライブラリを用いて作成した。モデル内部では、以下の 4 つの関数を回帰的に用いることで、分子構造を特徴ベクトルに変換した。

$$\text{Message function: } m_{vw}^t = M(s_v^t, s_w^t, e_{vw})$$

$$\text{Aggregate function: } m_v^t = \sum_{w \in \mathcal{N}_v} m_{vw}^t \cdot a_{vw}^t$$

$$\text{Update function: } s_v^{t+1} = U(s_v^t, m_v^t)$$

$$\text{Readout function: } s_{\text{mol}} = \sum_{v \in \mathcal{M}} s_v^T \cdot b_v$$

ここで  $M, U$  はニューラルネットワーク、 $s, e$  はノード (原子) およびエッジ (結合) の状態量、 $a, b$  はアテンションの重みを表している。最終的に得られた分子の状態量  $s_{\text{mol}}$  はベクトル量であり、その次元はハイパーパラメータの一つである。MPNN の事前学習には QM9 とよばれる構造最適化計算結果のデータベースを活用した。本データベースは、13 万分子以上の最適化構造 (原子の座標) および分子物性が収録されている。このデータベースの原子座標を入力値に、分子物性を出力値として事前学習を行った。事前学習を通じてモデルは分子構造の特徴をベクトルで出力することが可能となる。出力ベクトルの次元は 300 とした。

続いて、得られた分子の特徴ベクトルを入力値として物性推算を行った。物性推算には、沸点・融点・屈折率・密度・臨界温度・臨界圧力・臨界密度を用いた。各物性値は 155 ~ 1177 データを収集し、それらをランダムに訓練データ・検証データ・テストデータに分割した。訓練データはモデルのパラメータ更新に直接的に用い、検証データはモデル精度の汎用性を担保するために、損失関数のモニタリングにのみ用いた。テストデータは、モデルの訓練に一切関与させなかった。

また、実測値の収集のために、12-hydroxystearic acid (12HSA) を用いた超分子ゲルの物性測定を行った。ゲル生成において、自重で落下しない重量分率をゲル化率と定義し、ゲル生成可否の評価に用いた。超分子ゲルの機械的強度は、レオメーターによる応力依存測定により降伏応力を測定し評価した。また、ゲルに光を照射することで透過光量の測定も行った。

#### 4. 研究成果

Fig. 1 に、本研究で得られた MPNN が出力した分子の特徴ベクトルを用いた沸点、融点、臨界温度の推算誤差を示す。このモデルの訓練時には、訓練データ 80 点、検証データ 20 点を用い、沸点では 1177 点、融点では 832 点、臨界温度では 157 点のテストデータを用いた。データセットの分割を 10 回行い、その推算精度の平均値で評価したところ、いずれの物性においても既存のグループ寄与法である Joback-Reid 法を凌駕する精度で推算が可能であることが示された。学習に用いた学習データでは約 50 %程度の誤差の減少が確認されたうえ、学習に用いていないテストデータの推算精度も従来法と比較して 10～20%優れていた。従来法と比較して限られたデータ数であっても高精度の推算が可能であることから、実測が難しい物性の推算にも本手法が有用である可能性が示唆された。

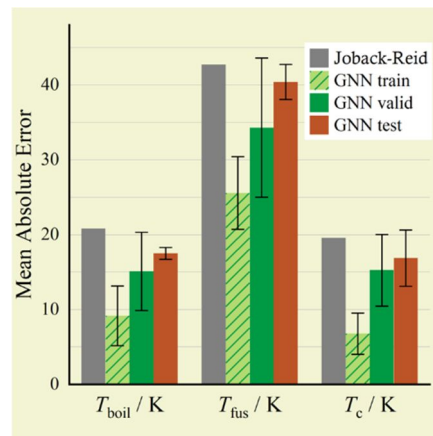


Fig. 1 MPNN の推算精度

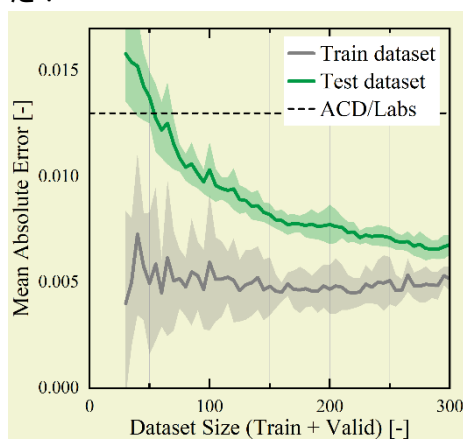


Fig. 2 データ数と推算精度

Fig. 2 には屈折率の推算において、訓練に用いたデータ数(訓練データと検証データの合計)と作成モデルの推算精度をまとめた。用いたデータの合計は 453 点である。訓練データの精度は、機械学習による推算の柔軟性により従来のソフトウェアを用いた推算精度を凌駕していた。加えて、テストデータの精度についても、データセットサイズが 50 点の時点で既存手法を凌駕していた。また、推算精度はデータセットサイズの上昇に伴い向上し、300 点のデータから学習を行う事で学習データと遜色のない精度で推算が可能であることが示された。同様の傾向は、いずれの物性でも確認され、本手法がわずかなデータから高精度な推算が可能であることが定量的に示された。

また、本手法は分子構造そのものを抽象的なベクトルに変換したのちに推算を行うため、未知の構造を有

する分子の物性推算に用いることができるという大きな利点を有する。一例として Fig. 3 には、アルデヒドを含まない学習データから作成した物性推算モデルを用いてアルデヒドの屈折率を推算した例を示す。モデルにとってアルデヒドは未知の構造を有しているにも関わらず、高い精度で推算が可能であることが示された。このような柔軟な推算は、従来のグループ寄与法のように、明示的に官能基の寄与を決定する手法では実現不可能であり、本手法の大きな利点であることが示された。

最後に、実際の物性への応用を指向し、低分子量ゲル化剤を用いたゲル作製およびその物性測定を行った。Fig. 4 には 12HSA ゲルの機械的強度を Hansen 溶解度パラメータを用いて整理した結果を示す。ゲル化剤分子と溶媒の相互作用によって生成するゲルの物性は、これらの 2 分子間のパラメータと強い相関があることが示された。今後は、上記の分子の特徴ベクトルを用いてゲル物性予測を行う事を目指す。

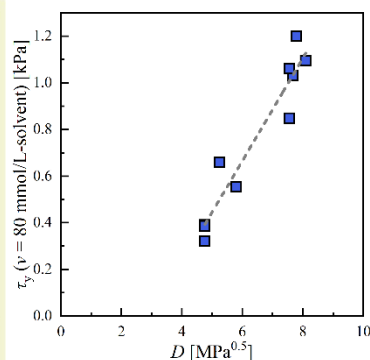
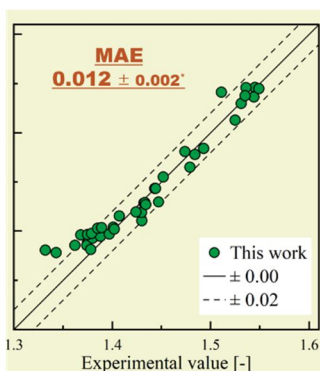
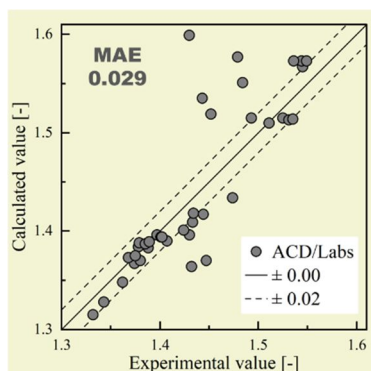


Fig. 3 アルデヒドの屈折率推算(左)既存手法, (右)本手法 Fig. 4 12HSA ゲルの機械的強度

#### 参考文献

- [1] F. A. Faber et al., *J. Chem. Theory Comput.*, **13** (2017) 5255-5264.

## 5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件／うち国際共著 0件／うちオープンアクセス 1件）

1. 著者名 Yuya Murakami, Taisei Uchiyama, Atsushi Shono	4. 巻 9
2. 論文標題 Correlation between physical properties of 12-hydroxystearic acid organogels and Hansen solubility parameters	5. 発行年 2023年
3. 雑誌名 Gels	6. 最初と最後の頁 314
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/gels9040314	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計3件（うち招待講演 2件／うち国際学会 0件）

1. 発表者名 村上 裕哉, 庄野 厚
2. 発表標題 Message Passing Neural Networkの転移学習を利用した分子三次元構造からの物性推算手法
3. 学会等名 分離技術年会2021（招待講演）
4. 発表年 2021年

1. 発表者名 宮下 純乃, 村上 裕哉, 松川 博亮, 大竹 勝人, 庄野 厚
2. 発表標題 低分子量ゲル化剤を利用した超分子ゲルの作製に関する研究
3. 学会等名 第25回化学工学会学生発表会
4. 発表年 2023年

1. 発表者名 村上 裕哉
2. 発表標題 畳込みグラフニューラルネットワークによる分子構造からの基礎物性の推算
3. 学会等名 化学工学会 第53回秋季大会（招待講演）
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------