

量子化学計算情報を記述子とした機械学習に基づく反応予測手法の開発

○藤波 美起登¹、清野 淳司²、中井 浩巳¹⁻⁴¹ 早稲田大学先進理工学研究科化学・生命化学専攻 (〒169-8555 東京都新宿区大久保 3-4-1)² 早稲田大学理工学研究所 (〒169-8555 東京都新宿区大久保 3-4-1)³ JST-CREST (〒332-0012 埼玉県川口市本町 4-1-8)⁴ 京大 ESICB (〒618-8520 京都府京都市西京区京都大学桂 京都大学ローム記念館 316 号室)

【緒言】これまで情報化学の分野において、コンピュータを用いて反応物から生成物を予測する手法が多数開発されてきた^[1]。しかし、予測精度の課題から商業的に普及するには至っていない。近年、機械学習を用いたシステム開発が諸分野で成果を挙げている。反応予測においても、化合物のグラフ構造情報を主な記述子とした機械学習により反応予測を行う、ReactionPredictor^[2]が開発され、その手法の有効性が示された。機械学習による高精度の予測には記述子が予測に適した情報を有することが不可欠である。そこで、本研究では、グラフ構造情報にかわり、非経験的な量子化学計算から得られる情報を記述子とする反応予測手法を開発し、その予測精度を検証した。

【理論と実装】Figure 1 にグラフ構造、量子化学計算から得られる記述子の相違を示す。グラフ構造から得られる記述子は、部分構造の有無により表現されるバイナリーである。一方、量子化学計算から得られる記述子は波動関数や電荷など分子の様々な情報を表す連続量である。量子化学計算情報は、記述子が包含する情報の多様化と記述子の連続量化により、少ない記述子数で反応を予測しうる情報を与えることが期待される。

本手法では反応のモデル化のために、局在化した軌道の情報を与える自然混成軌道を用いる。そして化学反応を「電子を供与する軌道から電子を受容する軌道への電子移動」と定義し、2段階の処理により反応予測を行う。第一段階では量子化学計算から得た記述子を用いて反応しうる軌道のスクリーニングを行う。第二段階ではスクリーニングを受けた軌道の組から、ランキング形式で相互作用しやすい軌道の組を予測し、生成物を得る。量子化学計算には、B3LYP の 6-31++G** (H-Ar) 、SDD (K-) を採用した。機械学習手法にはニューラルネットワークを、反応データベースには有機化学教科書^[3]内の 1110 の 2 電子移動反応を用いた。

【結果と考察】Table 1 に本手法 (Present) と ReactionPredictor において用いた反応データ数、記述子数を示す。Table 1 より、Present は ReactionPredictor と比べ、反応データ数、記述子数ともに少ない。Table 2 に各段階の予測精度を示す。Table 2 より、Present は ReactionPredictor と同等の精度のスクリーニングを実現した。また、84.6% の精度で正しい軌道相互作用を予測した。この予測精度は、反応データベース中のデータ数の増加に伴いさらに向上することが他の数値検証から示唆された。発表当日は学習する反応データ数と予測精度の関係、本手法によるラジカル反応などの予測結果も示し、本手法の性質、汎用性を議論する予定である。

参考文献

[1] W. A. Warr, *Mol. Inf.*, **33**, 469 (2014). [2] M. A. Kayala and P. Baldi, *J. Chem. Inf. Model.*, **52**, 2526 (2012). [3] M. Jones Jr. 著, 奈良坂 統一他監訳, ジョーンズ有機化学, 第 3 版, 東京化学同人 (2006).

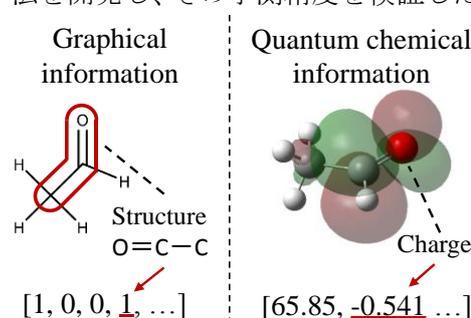


Fig 1. Illustration of descriptors in graphical and quantum chemical information

Table 1. Conditions of reaction prediction

		ReactionPredictor	Present
Number of reaction data		5551	1110
Number of descriptors	1st step	1500	24
	2nd step	1500	48

Table 2. Results of reaction prediction

		ReactionPredictor	Present
1st step	TNR ^{*1}	donor %	84.9
		acceptor %	68.8
	FNR ^{*2}	donor %	1.9
		acceptor %	1.3
2nd step	Accuracy of interaction prediction ^{*3}	98.5	84.6

*1 反応しない部位への正しいスクリーニングの割合

*2 反応する部位への誤ったスクリーニングの割合

*3 正しい軌道相互作用が上位 5 位以内に予測される割合