

令和 6 年 6 月 19 日現在

機関番号：13901

研究種目：基盤研究(C) (一般)

研究期間：2021～2023

課題番号：21K12027

研究課題名(和文) グラフニューラルネットワークによる有機遷移金属反応の機械学習

研究課題名(英文) Machine Learning of Organotransition Metal Reactions Using Graph Neural Networks

研究代表者

安田 耕二 (Koji, Yasuda)

名古屋大学・未来材料・システム研究所・准教授

研究者番号：70293686

交付決定額(研究期間全体)：(直接経費) 2,300,000円

研究成果の概要(和文)：多種類の有機遷移金属触媒反応を予測するグラフニューラルネットワーク(GNN)を構築した。素反応を学習対象に選び、電子則などを適切に強制することで、僅か数万個のパラメータを持つモデルでありながら、出発物と遷移金属+配位子を入力すると、最高で97.7%の精度で生成物を予測できた。学習データとして、原著論文などから出発物と生成物を収集し、素反応を推定、分解し、機械学習可能なデータベースを作った。GNNは各遷移金属の基質の好みを正しく予測した。GNNのノード潜在表現は、原子団の反応中の役割として理解できることが分かった。

研究成果の学術的意義や社会的意義

チタンやパラジウムなどの遷移金属を含む分子には、特定の化学結合を生成切断する触媒として、医薬品などの複雑な分子の合成に役立っている。有機化学者や理論化学者はこれまで、教科書や論文から反応パターンを学習し、新しい触媒反応を予想し実験して発見してきた。急速に発展しつつある深層学習を有機遷移金属反応に用いることで、プロ以上に有機遷移金属反応に詳しいAIを作る事を最終目標としている。適切なドメイン知識を用いることで、他研究の僅か1/1000のパラメータを用いて反応を高精度で予測できた。我々の機械学習モデルが量子化学計算を自律的に行い、反応探索を行える可能性が見えてきた。

研究成果の概要(英文)：A graph neural network (GNN) has been constructed to predict multiple types of organotransition metal catalytic reactions. By selecting elementary reactions as training targets and enforcing valence rules appropriately, we were able to predict products with up to 97.7% accuracy when starting materials and transition metal + ligands were input, even though the model has only a few tens of thousands of parameters. The GNN correctly predicted the substrate preference of each transition metal. The latent representation of a node in GNN was found to be understood as the role of the atomic group in the reaction.

研究分野：理論化学 機械学習

キーワード：深層学習 反応予測 グラフニューラルネットワーク 有機遷移金属

様式 C-19、F-19-1 (共通)

1. 研究開始当初の背景

チタンやパラジウムなどの遷移金属を含む分子には、特定の化学結合を生成切断する触媒として、医薬品などの複雑な分子の合成に役立っている。野依(2001年ノーベル化学賞)鈴木、根岸(2010年同賞)玉尾など多くの化学者が新反応を発見したが、これらは酸化的付加や移動挿入反応など、数種類の基本反応の組み合わせで理解できる。どの基本反応が生じるかは遷移金属の種類、周囲の配位子、反応分子などで変わり、多彩な遷移金属触媒を生む源になっている。機械学習、特にニューラルネットワークは、十分なデータがある(画像、翻訳)かデータが生成できる(囲碁)場合に、プロを超えた能力が発揮できる。ニューラルネットは近年グラフなどに応用範囲を広げ、複雑な有機分子の合成計画などが検討され始めた。

2. 研究の目的

この研究では、プロ以上に有機遷移金属反応に詳しいAIを作る事を最終目標としている。有機化学者や理論化学者はこれまで、教科書や論文から既存反応のパターンを学習し、新規反応を直感で予想し、実験やシミュレーションで発見してきた。彼らが行ってきた事を大規模にAIにさせ、彼らを超えるのが夢である。そのゴールに向かって3年間の研究期間では、既知反応を学習させ、それに基づき(量子化学計算は使わず)未知反応を予測できるようにしたい。有機遷移金属分子は普通の有機分子と異なりオクテット則に従わない難しさはあるが、原子を頂点、結合を辺とするグラフで分子構造式を表現できる。そこで反応させる分子をグラフニューラルネットに入力し、どの結合が生成又は開裂するかを、教師付きデータで学習させる。こうして教科書的な反応規則を教える事ができる。

3. 研究の方法

まず有機遷移金属反応の収集とデータクレンジングを行った。モデルの複雑さ(学習パラメータ数)にもよるが、機械学習には大量のデータが必要である。有機遷移金属反応のデータベースは公開、市販されていない。そこで、有機遷移金属化学の教科書や総説が引用している多数の原著論文と、US patent から収集された公開データから、実験で報告された反応(出発物と最終生成物)を集めた。有機遷移金属触媒反応は、出発物が幾つかの素反応を行って生成物となるため、原著論文や教科書を参考に各反応の機構を推定し、素反応に分解した。US patent データから約17万個、CAS(アメリカ化学会)から提供頂いた反応から約2万個の素反応が準備できた。素反応ごとに、種類(酸化的付加、還元的脱離など)や結合変化(できる/切れる結合)を記録し、酸化反応など単純なものを除き、学習用データセットとした。

4. 研究成果

出発物を与えると、結合変化を予想するグラフニューラルネットを作り、各素反応の出発物と生成物の対を与え、教師付きで学習を行った。予測精度の高いネットワーク構造を探した。結合の種類を考慮するいわゆるNNConvという畳み込みが有効だった。18電子則やHSAB原理などの基本ルールも学ばせる工夫をした(図1)。その結果、素反応のテストデータのうち、95%を正しく予測できた。金属により官能基や反応場所に好みがあるが、それが予測結果に正しく反映されていた(図2)。このグラフニューラルネットは、既知の有機遷移金属錯体の反応パターンを、とても良く「覚えている」と言える。素反応を学習対象に選んだこと、化学的に意味のあるネットワーク構造を選んだことが良かったと理解している。

既知の反応データベースを検索するのは異なり、未知だが類似した反応でも予測できる点の特筆される。また、ニューラルネットによる画像認識などと同様に、人が判別アルゴリズムや特徴量を設計する必要が無い。そのため、類似反応の検索や提案、計算機シミュレーションによる有望な反応の自動計算、に応用可能だと思われる。

深層学習モデルは推論過程を理解し難いことが多く、「説明可能性」の点で批判される。そこで我々

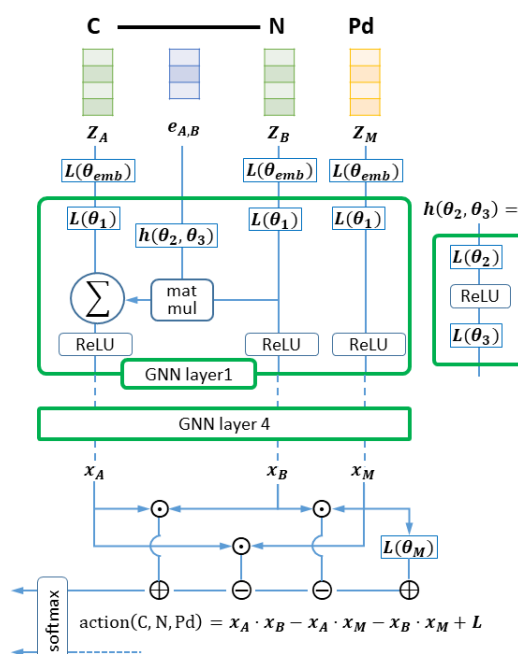


図1: グラフニューラルネットワークの構造。遷移金属以外では常にオクテット則を満たす。Lは遷移金属の空配座の影響を表す。

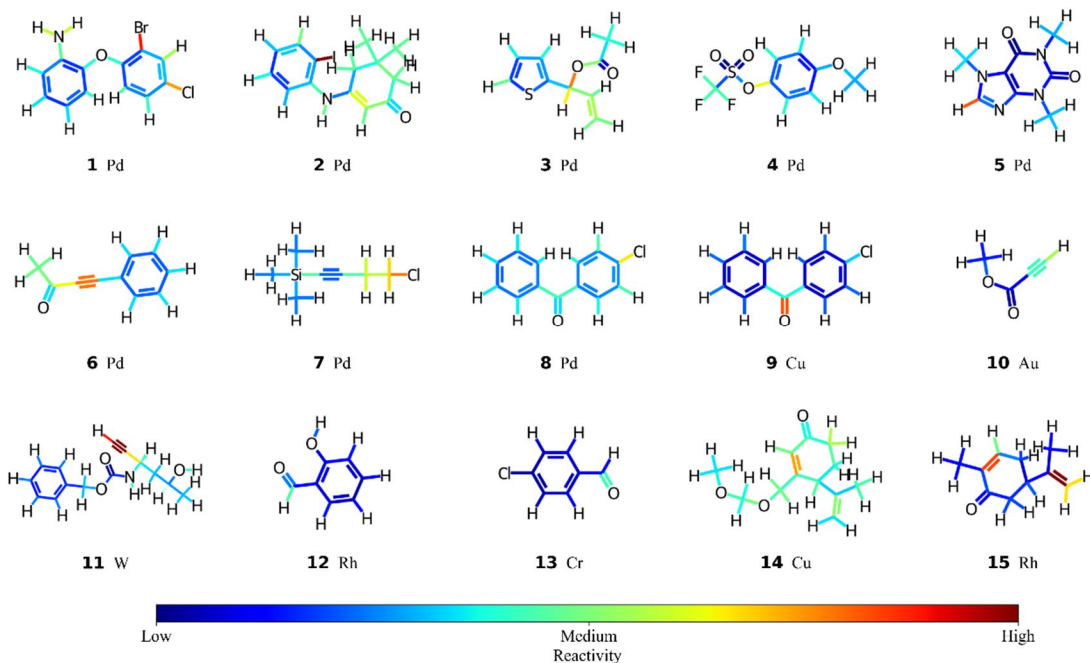


図2：酸化的付加反応の予測位置。基質と金属の好みを再現していることが分かる。

はグラフニューラルネットのノード潜在表現を解析した。適切な主成分分析を用いると、反応式中での役割に応じて部分構造がクラスターとして分類でき、潜在表現は理解可能だった。

学会、論文発表を通じて、実験化学者から「素反応を単純化しすぎている」との批判も受けた。素反応は反応の中間表現だと考えるとこれで構わないという意見もあるが、説明可能性の点では、実験化学者にも理解できるものがふさわしい。そこで主に以下の2点を改良した。

(1)触媒の配位子を約30種類に分類し、遷移金属+配位子を予測に用いた。

(2)酸化的付加、還元的脱離に加えて、素反応に配位を含めた。

その結果最高で97.7%の精度で生成物を予測できた。しかしdirecting groupや特別な配位子が重要なC-H活性化反応の予測精度は、その他反応より低く、将来の改善点として残った。

機械学習による化学反応予測は(遷移金属を含まない)純粋な有機反応の研究が殆どであり、またそのモデルは1000万個程度のパラメータを含む。他方、僅か数万個のパラメータを持つ我々のモデルは、より複雑(に見える)遷移金属触媒反応を高精度で予測できた。その理由は先行研究の多くがSMILESなどの線形表現で分子構造を表し、言語翻訳モデルを流用するという、化学者の理解とは程遠い不自然な方法だからと思われる。

有機遷移金属触媒反応は、出発物が幾つかの素反応を行って生成物となるが、通常の実験ではこの最終生成物しか観測しない。我々のモデルは教師付き学習をするため、素反応のデータベースを人が作る必要があった。この部分も自動化するため、作成した高精度の反応予測器を用いて、出発物と生成物のみを与え、多段階の素反応も予測する研究を行った。(出発物、生成物)に対して「迷路」探索を強化学習で行った。良いweightの初期値を使い、actor-critic法を用いると素反応を正しく予測できた。またhindsight experience replayを用いると安定に強化学習できるようになった。

化学反応の機械学習では、実験で得られたデータを一般に用いるため、入手やコストに問題がある。将来は、量子化学計算で反応データを追加するようになると思われるが、その際には素反応が学習単位となる。我々が実験データから作成した素反応データベースは、その出発点になり得る。更に、量子化学計算で反応データを自動生成するには、出発物の反応しそうな場所を選ぶ必要がある。我々の機械学習モデルはこのようなデータの自動生成にも適していると思われる。実際に実験をして予測を検証すること、Reaxysなど追加のデータセットを用いること、学習モデルを多数の実験化学者と競わせるテストを提案するアドバイスも受けたが、これらは予算で全く賄いきれなかった。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 0件）

1. 著者名 Motoji Sakai, Mitsunori Kaneshige, Koji Yasuda	4. 巻 45
2. 論文標題 Learning organo-transition metal catalyzed reactions by graph neural networks	5. 発行年 2023年
3. 雑誌名 J. Comput. Chem.	6. 最初と最後の頁 1-11
掲載論文のDOI (デジタルオブジェクト識別子) 10.1002/jcc.27243	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 酒井 基至, 金重 光典, 安田 耕二	4. 巻 21
2. 論文標題 グラフニューラルネットワークによる有機遷移金属反応の学習	5. 発行年 2022年
3. 雑誌名 J. Comput. Chem. Jpn.	6. 最初と最後の頁 126-128
掲載論文のDOI (デジタルオブジェクト識別子) 10.2477/jccj.2023-0012	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計5件（うち招待講演 0件/うち国際学会 2件）

1. 発表者名 酒井基至, 金重 光典, 安田耕二
2. 発表標題 グラフニューラルネットワークによる有機遷移金属反応の学習
3. 学会等名 日本コンピュータ化学会秋季年会
4. 発表年 2022年

1. 発表者名 Motoji Sakai, Mitsunori Kaneshige, Koji Yasuda
2. 発表標題 Learning Organo-Transition Metal Catalysis Using Graph Neural Networks
3. 学会等名 ICMaSS 2021 (国際学会)
4. 発表年 2021年

1. 発表者名 酒井基至, 金重光典, 安田耕二
2. 発表標題 グラフニューラルネットワークを活用した有機遷移金属触媒反応の学習
3. 学会等名 第15回分子科学討論会
4. 発表年 2021年

1. 発表者名 Motoji Sakai, Su Dehong, Koji Yasuda
2. 発表標題 Learning Organo-Transition Metal Catalyzed Reactions by Graph Neural Networks
3. 学会等名 TACC2023, Sapporo (国際学会)
4. 発表年 2023年

1. 発表者名 蘇 徳紅, 酒井 基至, 安田 耕二
2. 発表標題 グラフニューラルネットワークによる有機遷移金属反応の学習と深層強化学習の活用
3. 学会等名 日本コンピュータ化学会2023年 秋季年会
4. 発表年 2023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
---------	---------