

研究終了報告書

「量子化学と情報学との融合による次世代密度汎関数理論と均一系触媒における反応予測システムの開発」

研究期間：2017年10月～2021年3月

研究者：清野 淳司

1. 研究のねらい

本研究では、インフォマティクスと電子状態計算を融合させることで材料開発を加速させる二つの研究を遂行する。第一の研究は、電子状態計算手法として広く用いられてきた密度汎関数理論(DFT)の、最大の難題である厳密な汎関数の構築に挑戦する。第二の研究は、合成化学において必要不可欠である触媒を考慮した化学反応を予測するシステムを開発する。

第一の研究では、化合物の電子エネルギー項の中の運動エネルギー項に焦点を当て、満たすべき物理条件を暗に満たした膨大なデータから、精確な汎関数を構築することを目標とする。これによりインフォマティクスを利用した厳密な汎関数構築の可能性を示す。また構築された汎関数の導入により、従来の軌道を用いた Kohn-Sham DFT から、電子密度のみで表現可能な本来の DFT (OF-DFT) へと移行し、圧倒的に高速かつ大規模な電子状態を実現させるスキームを確立する。さらに上記の汎関数構築システムを相関エネルギー項に対しても適用できるように拡張を行い、任意の分子に対して定量的な議論ができる汎関数の構築を目指す。

第二の研究では、分子レベルでの反応の制御を可能にする均一系触媒に着目し、化学反応を直接支配する電子状態に関する情報とインフォマティクスを利用して、コンピュータにより化学反応を予測するシステムの開発を目指す。本研究ではさらに均一系触媒における化学反応予測のために重要な役割を果たし、かつ化学者が直感的にわかりやすい記述子の開発も行う。また化学反応は用いる触媒だけでなく、温度や濃度、溶媒などの種々の反応条件に実験結果が大きく左右される。そのため実験化学者と共同して、機械学習により様々な実験条件における実験データを効率的に収集できる手法の開発も行う。これらの研究を融合させることで、実際の実験に適用できる実用的な反応予測システムとすることを目指す。

2. 研究成果

(1) 概要

本研究の第一の目標である運動エネルギー汎関数の構築に関しては、本研究者はこれまでに微小体積(グリッド点)における記述子とエネルギー密度の関係をニューラルネットワークにより学習して汎関数を得る手法を確立した。本研究では、この手法を基に非局所型記述子の導入などの拡張を行うことで、従来の汎関数では困難であった化学結合の定性的な記述を達成した。さらに機械学習により構築した電子密度微分も導入することで、孤立系における OF-DFT 計算手法を確立し、電子密度の最適化ができることが確認された。

さらに相関エネルギーに拡張した。低レベル計算により得られた電子密度情報から、電子状態計算における Gold standard と評される CCSD(T)/CBS に基づく相関エネルギー密度を、機械学習により予測することで、定量的な電子状態計算を従来よりも数百倍高速にした。さらに凍結内殻近似に基づく機械学習型電子相関モデルを開発し、任意の元素への適用可能性を

示した。

第二の目標である触媒を考慮した化学反応予測システムに関しては、まず有機化合物同士の化学反応に対して、高い予測精度かつ電子状態情報を用いた直感的な解釈が可能な化学反応予測システムの基盤を確立した。

さらに均一系触媒における化学反応の予測性能向上や、より容易な解釈を目指して、電子状態に基づく記述子の開発を行った。具体的には、化学反応の強弱を表す原子間の結合エネルギーと、均一系触媒内の中心金属における電荷情報を量子化学計算に基づき役割の異なる三つの独立成分に分割する手法を開発した。

また均一系触媒における化学反応の様々な反応条件に対する実験データを効率よく収集するために、実験化学者と連携して、少ない実験データと多くの反応条件、または溶媒反応条件のみに対する機械学習を用いた最適化手法を開発した。その結果、実験化学者の知識や経験に依存せずに、効率的に最適な反応条件や溶媒を提案できた。

(2) 詳細

研究テーマ A「インフォマティクスを利用した DFT 汎関数の構築」

本研究者は、さきがけ研究採択前の同領域における特定課題調査において、DFT の Hohenberg-Kohn 定理に基づき、各グリッド点における記述子とエネルギー密度の関係をニューラルネットワークにより学習して汎関数を得る手法を確立した。これにより約 90 年以上に亘って開発されてきたすべての汎関数を凌駕する結果が得られた [J. Seino, et al. J. Chem. Phys. 148, 241705 (2018)]。本研究では、この手法を軸として実用的な精度の運動エネルギー汎関数を開発し、さらに高精度な相関エネルギーの予測に関する研究も遂行した。具体的な研究成果を以下に記す。

① 運動エネルギー汎関数の構築

本研究では、機械学習型運動エネルギー汎関数を用いて、様々な分子に対してポテンシャルエネルギー曲線(PEC)計算を行うことで、化学結合を正しく記述できるかを検証した。本手法の記述子として、各グリッド点における電子密度 ρ 、およびその 1 次、2 次、3 次勾配を用いた。また非局所的な効果を考慮するために各グリッド点に近い n 個の原子の中心からの距離の逆数も記述子として導入した。この結果、従来の汎関数で用いられてきた電子密度情報に加えて、高次の電子密度勾配と適切な距離の記述子を導入することで、KS-DFT の PEC を再現することが確認された(図 1)。これにより、本手法は OF-DFT における最大の問題の一つである化学結合の再現を解消できることが示された(研究成果 1)。

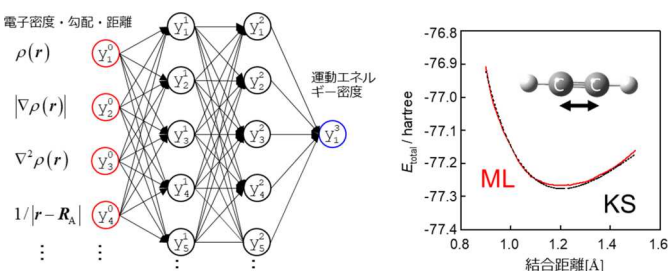


図 1. 機械学習型運動エネルギーに基づく PEC

さらに機械学習型運動エネルギー汎関数を導入した OF-DFT 計算手法を実現するために、二乗が電子密度と一致する関数を導入し、原子中心に置いた幾つかのガウス基底関数の線形結合を用いてエネルギーを表現する手法を開発・実装した。機械学習により構築された運

動エネルギー、およびその電子密度微分を組み込み、それぞれの原子・分子について最も低い全エネルギーを探索することで、最適な電子密度が得られる。この結果、本手法はこれまで提案されたどの運動エネルギー汎関数よりも、小さい誤差で KS-DFT の電子密度・運動エネルギーを再現することが確認された。

② 機械学習型電子相関モデルの構築

本研究では、簡便な電子状態計算手法である Hartree-Fock (HF) レベルと DZ などの小さい基底関数の結果から、Gold standard と評される CCSD(T)/CBS を再現する電子相関モデルを機械学習により構築した。記述子としてグリッドごとの電子密度やその勾配、運動エネルギー密度、HF 交換エネルギー密度を、目的変数として本研究者が開発した複合法による CBS 極限 [J. Seino et al. J. Comput. Chem. 37, 1517 (2016)] での、グリッドエネルギー密度解析による CCSD(T) 相関エネルギー密度を用いた。その結果、参照となる CCSD(T)/CBS を化学的精度 (誤差 1 kcal/mol 以内) で予測できることが確認された。また例えばメタン分子の計算時間において、通常の CCSD(T)/CBS で約 1 万 8 千秒かかる計算が、電子相関モデルでは約 50 秒と圧倒的に高速な計算が達成された (研究成果 2)。

さらに第 3 周期以降の元素にも適用するために、化学物性を主に支配する価電子軌道のみに着目し、内殻軌道を見捨てる凍結内殻近似を導入して記述子・目的変数を算出し、新たな電子相関モデルを構築した。この結果、全電子の場合と同様の精度で適用可能であることが確認された。実際に CH_3Br 分子の目的変数の算出に要する計算時間は、全電子計算の 177.8 時間から 1.9 時間となり、約 90 倍程度高速化された。

研究テーマ B「均一系触媒における化学反応予測システムの開発」

本研究では、均一系触媒における化学反応の可否を、機械学習により電子レベルで決定する反応予測システムを開発した。具体的には、反応予測システム基盤の構築、化学反応の解釈を容易にするための記述子の開発、および反応条件の違いによる微細な実験結果の変化を考慮した効率的なデータ取得手法の開発を行った。

① 機械学習を利用した反応予測システム基盤の確立

本研究者はこれまでに、電子状態に関する情報を記述子とした機械学習に基づき、有機化合物の反応予測を行う手法を開発した。本研究では、さらにペアワイズ型学習に変更することでデータの正例と負例の数のギャップの問題を解消し、ランダムフォレストを導入することで予測精度を向上させた。この結果、主として電子状態の情報を記述子とした場合でも、従来のトポジカルな記述子と同等の結果を与えることを確認した。さらに本手法の記述子解析により、電荷 (Q) や Fukui 指数 (f) が最も寄与が大きく、化学者の直感と合致する結果が得られた (図 2)。これにより高い精度で予測可能かつ直感的な解釈ができる化学反応予測システムの基盤が確立された (研究成果 3)。

② 反応の解釈を容易にする記述子の開発

化学結合の強弱の情報は化学反応が起こるかどうかの判定に重要な役割を果たす。本研究では量子化学計算に基づく結合エネルギー分割と LASSO を組み合わせることで、共有結合・イオン結合・水素結合などの結合の強弱を問わず、結合エネルギーを算出できる新たな

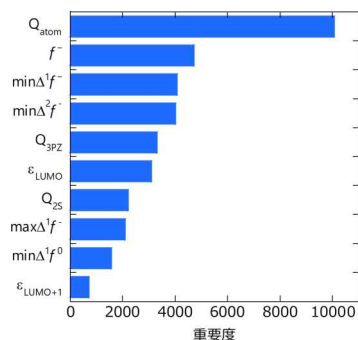


図 2. 反応予測システムの記述子解析

方法を開発した。その結果、共有結合・イオン結合を含んだ種々の化合物に対して、実験でも確認できる結合解離エネルギー(BDE)の計算値と良い相関が得られた。本手法は BDE の算出が難しいベンゼン環など、あらゆる分子に対しても適用できる。

また均一系触媒における化学反応において、中心金属における電子状態は反応の可否を決める重要な情報となりうる。遷移金属における幾何構造の推定や反応性の解釈などを示す経験的な指標として、古くから形式酸化数が用いられてきた。しかし、形式酸化数は必ずしも実際の電子状態を反映しない。そこで本研究では、量子化学計算に基づき遷移金属の酸化数を再定義することで、電荷情報から有用な成分へと分割する方法を提案した。その結果、典型的な遷移金属錯体に対して、従来の経験的な手法と等価な酸化数を与えた。さらにその他の成分も化学結合形式や、配位子環境下の遷移金属の状態を反映した記述子として利用できることも確認された。

③ 実験・計算・インフォマティクスが融合した実験条件最適化手法の開発

実験化学において、反応条件の最適化は不可欠なプロセスである。この効率化により、反応条件を考慮した反応予測システムを構築するための有効なデータを多く取得することが可能である。本研究ではテトラシラビシクロブタンの合成反応に着目し、温度や溶媒など 9 つの実験条件について最大の目的値を与える条件を、機械学習により効率的に探索する手法を開発した(図 3)。機械学習手法にはカーネルリッジ回帰を用いた。実験データは東北大学の岩本武明研究室より提供いただいた。この結果、実験では最大の目的値を得るのに 50 回かかるのに対し、機械学習を用いる手法は平均 11.7 回と効率化されることが確認された。特に機械学習を用いる最初の段階で、データ間の距離が遠いデータ群を用いることが重要であることが示された。

反応条件の最適化において、条件の一つである溶媒は収率などに大きく影響を与えるため、特に優先的に検討される。本研究ではパラジウム触媒を用いたエステルの脱カルボニルエーテル化反応に着目し、最適な溶媒の予測を行った。溶媒の記述子として、溶媒のデータベースや量子化学計算から得られた 17 種の物性値を用いた。目的値として、早稲田大学の山口潤一郎研究室から提供いただいた実験収率を用いた。機械学習として、LASSO による回帰および、k-means 法によるクラスタリングを行った。この結果から、クラスタリングによる分類は実験収率の大小と相関関係があり、さらに回帰による予測も実験収率を定性的に再現した。回帰とクラスタリングの両面から高収率を与える溶媒を提案できることが確認された。

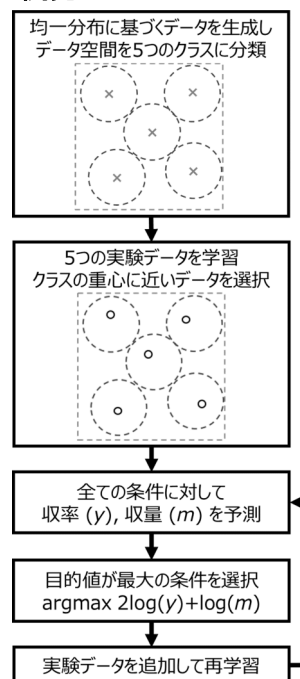


図 3. 実験条件最適化のアルゴリズム

3. 今後の展開

研究テーマ A では、電子密度とエネルギー密度の関係を機械学習で繋ぎ、汎関数を構築するシステムにより、運動/相関エネルギーに対する高度な汎関数を得ることができた。現状まではある程度限定された分子系に対する基礎的な検証が軸であったため、今後は汎用性を向

上させることで実用段階まで発展させ、誰でも使用できるように公開する予定である。そのために GDB-9 や PubChemQC のような大規模分子データベースから自動的に分子構造を抽出し、密度データベースを構築することが不可欠となる。さらにデータベースとオンライン機械学習と連結することで生成されたデータを常時学習し、汎用性を向上し続けるシステムとする。本システムの完成により、従来の DFT では計算できなかった分子サイズ(数十万原子以上)の全電子状態を定量的な議論が可能な精度で得られる手法となることが期待できる。

研究テーマ B では、均一系触媒における化学反応の可否を予測するための基盤システム、化学反応を表現する記述子、有効なデータを生成するための手法を開発した。しかし、実際に実験化学者が使用するためには、化学反応が起きるか否かだけでなく、実験条件を変化させたことによる収率などの量的な予測も重要である。この実現のために実験化学者との共同研究を拡大し、多くのデータを集約させたデータベースを作成していく必要がある。同時に研究テーマ A の手法を用いることで、実際の実験環境下の化学反応を再現した電子状態計算から反応予測に関する定量的なデータを取得することも検討していく。

4. 自己評価

研究テーマ A における、インフォマティクスを利用して実用的な精度の運動エネルギー汎関数を構築し、OF-DFT に基づく計算手法を確立するという当初の目的は概ね達成した。さらに申請時の計画にはなかった相関エネルギーまで拡張し、定量的な電子状態計算の高速化まで到達した。上記の基礎的な検証はほぼ終了し、実用性を目指したオンライン機械学習への拡張も現在進行形で進んでいる。本研究課題終了時までには間に合わないものの公開に向けた準備も順調である。

研究テーマ B では、有機化合物同士の化学反応予測や Pd 均一系触媒の酸化的付加反応など、限定した系に対する化学反応の可否を予測することは可能になった。しかし、領域内会議や実験化学者との議論を通して、実際に使用するシステムとするには実験収率などの量的な予測が重要であり、現状のシステムでは不十分であることを認識した。そこで、計算による量的なデータ収集を想定して研究テーマ A に少し比重を置いて研究を進めた。さらに、実験における量的なデータを収集するために、2 つの実験研究室との共同研究を実施した。この共同研究を通して、当初の計画にはなかった効率的に実験条件最適化を行う手法を開発できた。また、本さきがけ研究を通して多くの大学・企業の研究者と化学×インフォマティクスに関する議論ができたのも収穫であった。今後もこれらの研究を発展させることにより共同研究の輪を広げて、化学におけるインフォマティクスの重要性を広めていきたい。

研究テーマ A における定量的な高速電子状態計算により、計算のみから革新的な機能性材料となりうる化合物を選定し、研究テーマ B で開発した手法により、選定された化合物を得るための合成指針を得ることができる。このように両研究テーマの統合により、材料開発を包括的に加速させるシステムとなることが期待できる。現状において、両テーマともまだ途上段階と言えるが、早期に完遂させることで、研究成果の波及を目指していきたい。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数:12件

1. J. Seino, R. Kageyama, M. Fujinami, Y. Ikabata, H. Nakai, “Semi-local machine-learned kinetic energy density functional demonstrating smooth potential energy curves”, Chem. Phys. Lett. 2019, 734, 136732-1-6.

本研究では、機械学習型運動エネルギー汎関数を用いて、様々な分子に対してポテンシャルエネルギー曲線(PEC)計算を行うことで、化学結合を正しく記述できるかを検証した。この結果、従来の汎関数で用いられてきた電子密度情報に加えて、高次の電子密度勾配と適切な距離の記述子を導入することで、KS-DFT の PEC を再現することが確認された。これにより、本手法はOF-DFTにおける最大の問題の一つである化学結合の再現を解消できることが示された。

2. T. Nudejima, Y. Ikabata, J. Seino, T. Yoshikawa, H. Nakai, “Machine-learned electron correlation model based on correlation energy density at complete basis set limit”, J. Chem. Phys. 2019, 151, 024104-1-12.

本研究では、簡便な電子状態計算手法である Hartree-Fock(HF)レベルと DZ などの小さい基底関数の結果から、Gold standard と評される CCSD(T)/CBS を再現する電子相関モデルを、1.の論文における手法を拡張することにより構築した。その結果、参照となる CCSD(T)/CBS を化学的精度(誤差 1kcal/mol 以内)で予測できることが確認された。また例えばメタン分子の計算時間において、通常の CCSD(T)/CBS で約 1 万 8 千秒かかる計算が、電子相関モデルでは約 50 秒と圧倒的に高速な計算が達成された。

3. M. Fujinami, J. Seino, H. Nakai, “Quantum chemical reaction prediction method based on machine learning”, Bull. Chem. Soc. Jpn. 2020, 93, 685-693, Selected Paper.

本研究者はこれまでに、電子状態に関する情報を記述子とした機械学習に基づき、有機化合物の反応予測を行う手法を開発した。本研究では、さらにペアワイズ型学習に変更することでデータの正例と負例の数のギャップの問題を解消し、ランダムフォレストを導入することで予測精度を向上させた。この結果、主として電子状態の情報を記述子とした場合でも、従来のトポロジカルな記述子と同等の結果を与えることを確認したこれにより高い精度で予測可能かつ直感的な解釈ができる化学反応予測システムの基盤が確立された。

(2)特許出願

研究期間累積件数: 0 件

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

主要な学会発表

1. J. Seino, R. Kageyama, M. Fujinami, Y. Ikabata, H. Nakai, “Orbital-Free Density Functional Theory Calculation Combined with Semi-Local Machine-Learned Kinetic Energy Density Functional”, The 2nd Global Forum on Advanced Materials and Technologies for Sustainable Development (GFMAT-2), Toronto(Canada), 2019 年 7 月 21 日-26 日, Invited.
2. 清野淳司、「インフォマティクスを導入した密度汎関数理論の開発」、第 29 回格子欠陥フォーラム「格子欠陥とマテリアルズ・インフォマティクス」、豊田工業大学、2019 年 9 月 14 日、15 日、招待講演.

3. 清野淳司、「人工知能技術がもたらす理論化学の深化と進化」、日本化学会第 101 春季年会(2021) 併催シンポジウム「30 年後の夢をかなえる理論化学」、オンライン、2021 年 3 月 20 日、招待講演.

著作物

4. 清野淳司、中井浩巳、「インフォマティクスとの融合による理論化学研究」、化学工業(特集「マテリアルズ・インフォマティクスの未来展望」)、化学工業社、2018, 69, 53-58.
5. 藤波美起登、清野淳司、中井浩巳、「人工知能を用いた化学反応の予測と反応条件の最適化」、マテリアルズ・インフォマティクスによる材料開発と活用集・第 8 章、技術情報協会、2019, 379-384.