

令和 3 年 6 月 11 日現在

機関番号：32689

研究種目：若手研究

研究期間：2018～2020

課題番号：18K14184

研究課題名(和文)インフォマティクスを用いたユニバーサル交換相関汎関数の構築

研究課題名(英文)Construction of a universal exchange-correlation functional using informatics technology

研究代表者

五十幡 康弘 (IKABATA, Yasuhiro)

早稲田大学・理工学術院・次席研究員(研究院講師)

研究者番号：10728166

交付決定額(研究期間全体):(直接経費) 3,100,000円

研究成果の概要(和文):本研究課題では、Hartree-Fock計算で得た電子密度等からCCSD(T)法の完全基底関数で得られる相関エネルギー密度を予測する機械学習型電子相関(ML-EC)モデルを開発した。グリッドエネルギー密度解析と複合法により相関エネルギー密度を定義し、機械学習の目的変数とした。機械学習のモデルとしてニューラルネットワークを採用し、相関エネルギーや化学物性の高精度な再現に成功した。さらに凍結内殻近似の適用により目的変数の計算コストが削減され、重元素を含む系の学習が可能となった。これらの成果により、高速かつ汎用的な電子相関モデルの構築への道が拓けた。

研究成果の学術的意義や社会的意義

量子化学計算は化学における研究開発の多くで普及しているが、より高精度かつ高速な計算手法の開発は依然として重要である。近年、計算手法の開発に機械学習を用いるアプローチが注目されており、本研究課題で開発した機械学習型電子相関(ML-EC)モデルはその一つとして位置づけられる。相関エネルギー密度を電子密度の汎関数としてフィッティングする方法は以前から報告されていたが、ニューラルネットワークの使用により精度が飛躍的に向上した。本研究課題の成果に基づき汎用性の高い機械学習モデルを構築し公開することで、量子化学計算を用いた研究開発の促進に貢献することが期待される。

研究成果の概要(英文): In this project, we have developed the machine-learned electron correlation (ML-EC) model, which predicts the correlation energy density at the complete basis set limit of the CCSD(T) method from electron density and other density variables obtained by the Hartree-Fock calculation. The correlation energy density was defined by grid-based energy density analysis and composite method and adopted as the objective variable of machine learning. The neural network successfully reproduced the correlation energy and chemical properties with high accuracy. The frozen core approximation significantly reduced the computational cost of the objective variable, which enabled machine learning of heavy-element systems. These results paved the way for the construction of a general-purpose correlation model.

研究分野：理論化学、量子化学、計算化学

キーワード：インフォマティクス 機械学習 電子相関 密度汎関数理論 ニューラルネットワーク

様式 C-19、F-19-1、Z-19 (共通)

1. 研究開始当初の背景

Kohn-Sham 密度汎関数理論 (KS-DFT) は、今日の量子化学計算の多くで利用される電子状態理論である。KS-DFT の計算精度は交換相関汎関数の近似に左右され、これまでに多数の近似汎関数が開発されてきた。Perdew は汎関数の系統的な高精度化を旧約聖書の「Jacob の梯子」になぞらえて説明した。これは、地上 (交換相関項がない Hartree 理論) と天国 (化学的精度、熱力学量の誤差が 1 kcal/mol 以下の理論) をつなぐ 5 段からなる梯子である。1 段目の汎関数は電子密度のみに依存し、2 段目では密度勾配にも依存し、3 段目では運動エネルギー密度にも依存する。4 段目は Hartree-Fock 交換を混成した汎関数が該当し、B3LYP など量子化学計算で用いられる汎関数の多くが該当する。5 段目は非占有軌道に依存する汎関数が該当し、二重混成汎関数などが提案されている。

Jacob の梯子に沿った汎関数の開発によって DFT 計算の精度は向上してきたが、いかなる系、物理量に対しても高精度な汎関数 (ユニバーサル汎関数) には程遠い。計算対象となる系によって適切な汎関数が異なるため、実験値を再現する汎関数や、ベンチマーク論文で誤差が小さいと報告された汎関数が選択される。このような背景から、従来型の物理条件に基づく定式化と化学物性に対するパラメータフィッティングによる汎関数の開発とは異なるアプローチが求められていた。

2. 研究の目的

本研究課題では、インフォマティクスの技術を用いて大量のデータに適合した汎関数を構築し、ユニバーサル汎関数を提示することを目的とした。インフォマティクスの技術として、群知能を用いたパラメータ最適化と、機械学習における教師あり学習に注目した。

3. 研究の方法

本研究課題の初年度では、経験的パラメータを多数含んだ交換相関汎関数の表式を考案し、群知能を用いてパラメータの最適化を行う予定であった。しかし、検討の結果、この方法で従来の精度を超える汎関数を構築するのは容易でないと判断した。一方、研究代表者らは軌道非依存 DFT に対して従来の精度を上回る運動エネルギー汎関数を、ニューラルネットワークを用いて構築することに成功していた (*J. Chem. Phys.* **148**, 241705 (2018))。そこで、本研究課題の 2 年目以降に予定していた機械学習を用いた汎関数の構築を、運動エネルギー汎関数の構築と同様の手法で初年度から実施することにした。

Gauss 基底を用いた量子化学計算では、Hartree-Fock 交換エネルギーは容易に計算できる一方で、電子相関エネルギーの高精度計算には多大な計算コストを要するため、相関汎関数のみを構築する方針に変更した。また、KS-DFT の枠組みでは相互作用のない参照系の運動エネルギーと厳密な運動エネルギーの差が相関汎関数に含まれるが、このエネルギー差を任意の分子について計算するのは困難である。そこで、波動関数理論に基づく高精度計算から電子相関エネルギーの汎関数を機械学習型電子相関 (ML-EC) モデルとして構築し、モデルを使用する際には Hartree-Fock 計算から得られる記述子を用いて post-Hartree-Fock 的に相関エネルギーを計算することにした。

機械学習に必要な記述子と目的変数を計算するコードは、GAMESS プログラムに実装した。機械学習のモデルとしてニューラルネットワークを採用し、ライブラリには Chainer および PyTorch を使用した。構築したモデルを用いた相関エネルギーおよび全エネルギーの評価には、GAMESS および RAQET プログラム (*J. Comput. Chem.* **39**, 2333 (2018)) を用いた。プログラムのコンパイル、機械学習、量子化学計算などは研究代表者が所属する研究室のワークステーションおよび自然科学研究機構のスーパーコンピュータにて実行した。

4. 研究成果

(1) 機械学習型電子相関モデルの構築と数値検証

DFT における交換相関汎関数の精度は、CCSD(T)法の完全基底 (CBS) 極限を参照値として評価される。そこで、本研究では CCSD(T)/CBS レベルの相関エネルギー密度を機械学習における目的変数とした。目的変数をエネルギー密度にすることで、少数の分子から機械学習を行うために十分な量のデータを得ることができる。MP2 法や CCSD 法に対しては、グリッドエネルギー密度解析による相関エネルギー密度 (*J. Comput. Chem.* **29**, 1555 (2008)) が提案されており、本研究ではこれを CCSD(T)法に拡張した。さらに、相関エネルギーの CBS 極限を見積る手法である外挿法 (*J. Chem. Phys.* **106**, 9639 (1997)) や複合法 (*J. Comput. Chem.* **37**, 2304 (2016)) を、エネルギー密度に適用して CCSD(T)/CBS レベルの相関エネルギー密度を見積った。外挿法と複合法による CBS 極限と quintuple-zeta 基底関数系の相関エネルギー密度は非常に近く、計算コストの小さな複合法が目的変数の計算に有用であることがわかった。

H, C, N, O 原子からなる 15 種類の小分子をトレーニングセットに設定し、31,771 点のグリッド点における記述子と目的変数を計算し、ML-EC モデルを構築した。記述子として電子密度、

密度勾配、運動エネルギー密度、Hartree-Fock 交換エネルギー密度、そして非整数占有数に基づく電子密度を採用した。構築した ML-EC モデルは Jacob の梯子における 5 段目の汎関数に相当する。効率的な相関エネルギーの見積りを可能とするために、記述子は double-zeta 基底関数系による Hartree-Fock 計算で得ることとした。構築した ML-EC モデルは、48 種類の分子で構成されるテストセットに対して CCSD(T)/CBS レベルの相関エネルギーを平均誤差 -0.7 mHartree、平均絶対誤差 1.8 mHartree で再現した。12 種類の反応エネルギーの精度を検証したところ、ML-EC モデルは既存の汎関数 73 種による DFT 計算を上回る精度を与えた (表 1)。

表 1. 12 種類の化学反応に対する反応エネルギーの平均絶対誤差 (MAD) および最大絶対誤差 (MaxAD). 数値の単位は kcal/mol. CCSD(T)/CBS の値を参照値とした.

	MAD		MaxAD
ML-EC	0.892	ML-EC	1.729
M06-2X	1.423	B3LYP	3.645
PW6B95	1.585	X3LYP	3.688
X3LYP	1.608	M11	3.991
MN15	1.614	M08-HX	4.283
M06	1.637	M06	4.291

続いて ML-EC モデルの汎用性を調査するために系統的な数値検証を行った。GMTKN55 データベース (*Phys. Chem. Chem. Phys.* **19**, 32184 (2017)) に注目し、第 3 周期までの閉殻系を対象として化学物性の計算精度を検証した。トレーニングセットは GMTKN55 データベースから抽出した原子、小規模な分子、その二量体で構成した。反応エネルギーや反応障壁については、構築したモデルをより大きな系に適用しても破綻せず妥当な精度の計算結果が得られた。分子間相互作用については、分散力が過小評価される傾向がある一方、静電相互作用や水素結合は問題なく記述された。また、構築したモデルにおけるパラメータ数は非常に多い (十万のオーダー) にも関わらず、トレーニングセットに含まれる元素以外を含む系に適用しない限り、数値的に安定した挙動を示した。

(2) 凍結内殻近似に基づく機械学習型電子相関モデルの構築と数値検証

ML-EC モデルを構築するために行う CCSD(T)/CBS レベルの相関エネルギー密度の計算は、内殻電子が多い高周期元素を含む系では計算コスト (計算時間およびメモリ使用量) が大きいため困難である。内殻電子の化学物性への影響は小さいので、CCSD(T)法などの波動関数理論に基づく電子相関エネルギーの計算では、凍結内殻近似によって内殻電子の軌道を考慮せず計算コストを削減するのが一般的である。そこで、相関エネルギー密度の計算に凍結内殻近似を適用することで計算コストを削減し、ML-EC モデルを構築するための目的変数をより広範な系から収集することを目指した。予想された通り、内殻軌道の数が多い第 3 周期や第 4 周期の元素を含む分子において、相関エネルギー密度に要する計算コストの減少は顕著であった。

H, C, N, O, F, S, Cl 原子からなる 30 種類の小分子をトレーニングセットに設定し、58,405 点のグリッド点における記述子と目的変数を計算し、ML-EC モデルを構築した。記述子として電子密度、密度勾配、運動エネルギー密度、Hartree-Fock 交換エネルギー密度を採用した。構築したモデルは Jacob の梯子における 4 段目の汎関数に相当する。15 種類の反応エネルギーの精度を検証したところ、ML-EC モデルは既存の汎関数 71 種による DFT 計算を上回る精度を与えた (表 2)。

表 2. 15 種類の化学反応に対する反応エネルギーの平均絶対誤差 (MAD) および最大絶対誤差 (MaxAD). 数値の単位は kcal/mol. CCSD(T)/CBS の値を参照値とした.

	MAD		MaxAD
ML-EC	1.226	ML-EC	2.997
M06-2X	1.260	M06-2X	3.488
M06	1.855	M06	4.592
X3LYP	2.002	B97-3	5.184
B3LYP	2.046	X3LYP	5.234
B97-3	2.093	BPBE	5.413

(3) 機械学習型電子相関モデルと RAQET プログラムの接続

ML-EC モデルを用いて CCSD(T)/CBS レベルの全電子エネルギーを近似的に見積るには、3 種類の基底関数 (例えば cc-pVDZ, cc-pVTZ, cc-pVQZ) による Hartree-Fock 計算を行い、小規模な基底関数 (cc-pVDZ) を用いて計算した記述子を用いて相関エネルギーを計算する。図 1 に水分子クラスターのベンチマークセットである WATER27 について、RAQET プログラムによる Hartree-Fock 計算、記述子の計算、および ML-EC モデルによる相関エネルギーの計算に要した

wall time を示す。RAQET プログラムには、ML-EC モデルの記述子のうち最も計算時間を要する Hartree-Fock 交換エネルギー密度を効率的に計算するためのグリッドルーチンやスクリーニング法 (*J. Chem. Theory Comput.* **15**, 4745 (2019)) が実装されている。これにより、記述子の計算 (図 1 における Descriptor) を効率的に行うことができる。また、相関エネルギーの計算時間 (図 1 における Correlation) は分子数に関してほぼ線形であり、ML-EC モデルにより相関エネルギーを効率的に評価できることが示された。

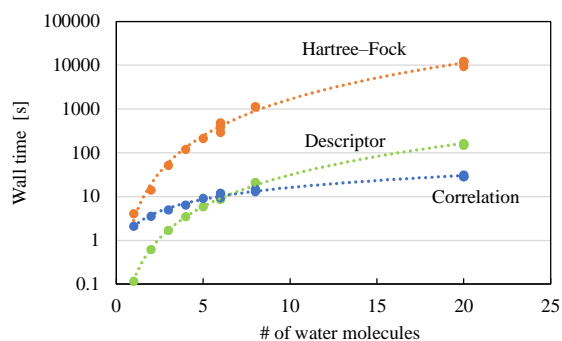


図 1 . WATER27 サブセットに対して ML-EC モデルを用いた高精度電子エネルギー計算に要した計算時間 . CPU は Intel Xeon Gold 6148 , 40 CPU コアを使用 .

5. 主な発表論文等

〔雑誌論文〕 計20件（うち査読付論文 20件 / うち国際共著 5件 / うちオープンアクセス 6件）

1. 著者名 Yasuhiro Ikabata, Ryo Fujisawa, Junji Seino, Takeshi Yoshikawa, Hiromi Nakai	4. 巻 153
2. 論文標題 Machine-learned electron correlation model based on frozen core approximation	5. 発行年 2020年
3. 雑誌名 The Journal of Chemical Physics	6. 最初と最後の頁 184108
掲載論文のDOI (デジタルオブジェクト識別子) 10.1063/5.0021281	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 五十幡 康弘	4. 巻 2
2. 論文標題 機械学習を用いた運動エネルギー汎関数と電子相関モデルの構築	5. 発行年 2020年
3. 雑誌名 理論化学会誌「フロンティア」	6. 最初と最後の頁 72 ~ 80
掲載論文のDOI (デジタルオブジェクト識別子) なし	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 五十幡 康弘, 吉川 武司, 中井 浩巳	4. 巻 18
2. 論文標題 相対論的量子化学計算プログラムRAQETの公開	5. 発行年 2019年
3. 雑誌名 Journal of Computer Chemistry, Japan	6. 最初と最後の頁 A6 ~ A11
掲載論文のDOI (デジタルオブジェクト識別子) 10.2477/jccj.2019-0022	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 -
1. 著者名 Toni M. Maier, Yasuhiro Ikabata, Hiromi Nakai	4. 巻 15
2. 論文標題 Efficient Semi-Numerical Implementation of Relativistic Exact Exchange within the Infinite-Order Two-Component Method Using a Modified Chain-of-Spheres Method	5. 発行年 2019年
3. 雑誌名 Journal of Chemical Theory and Computation	6. 最初と最後の頁 4745 ~ 4763
掲載論文のDOI (デジタルオブジェクト識別子) 10.1021/acs.jctc.9b00228	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 Takuro Nudjima, Yasuhiro Ikabata, Junji Seino, Takeshi Yoshikawa, Hiromi Nakai	4. 巻 151
2. 論文標題 Machine-learned electron correlation model based on correlation energy density at complete basis set limit	5. 発行年 2019年
3. 雑誌名 The Journal of Chemical Physics	6. 最初と最後の頁 24104
掲載論文のDOI (デジタルオブジェクト識別子) 10.1063/1.5100165	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 Junji Seino, Ryo Kageyama, Mikito Fujinami, Yasuhiro Ikabata, Hiromi Nakai	4. 巻 148
2. 論文標題 Semi-local machine-learned kinetic energy density functional with third-order gradients of electron density	5. 発行年 2018年
3. 雑誌名 The Journal of Chemical Physics	6. 最初と最後の頁 241705
掲載論文のDOI (デジタルオブジェクト識別子) 10.1063/1.5007230	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39件 (うち招待講演 1件 / うち国際学会 12件)

1. 発表者名 Yasuhiro Ikabata, Takuro Nudjima, Junji Seino, Takeshi Yoshikawa, Hiromi Nakai
2. 発表標題 Machine-Learned Correlation Model for Accurate and Efficient Computation of Correlation Energy
3. 学会等名 The 5th International Conference on Molecular Simulation (ICMS2019) (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Yasuhiro Ikabata, Takuro Nudjima, Junji Seino, Takeshi Yoshikawa, Hiromi Nakai
2. 発表標題 Machine-learned electron correlation model for accurate reproduction of correlation energy at the basis-set limit
3. 学会等名 The Ninth Conference of the Asia-Pacific Association of Theoretical and Computational Chemists (APATCC 2019) (国際学会)
4. 発表年 2019年

1. 発表者名 Yasuhiro Ikabata, Takuro Nudjima, Junji Seino, Takeshi Yoshikawa, Hiromi Nakai
2. 発表標題 Machine-Learned Electron Correlation Model in the Form of Density Functional
3. 学会等名 18th International Conference on Density-Functional Theory and its Applications (DFT2019) (国際学会)
4. 発表年 2019年

1. 発表者名 藤澤遼, 五十幡康弘, 清野淳司, 吉川武司, 中井浩巳
2. 発表標題 機械学習を用いた価電子の相関エネルギー予測
3. 学会等名 第10回CSJ化学フェスタ2020
4. 発表年 2020年

1. 発表者名 藤澤遼, 五十幡康弘, 清野淳司, 吉川武司, 中井浩巳
2. 発表標題 機械学習型電子相関モデルの開発: 凍結内殻近似の適用
3. 学会等名 分子科学オンライン討論会
4. 発表年 2020年

1. 発表者名 Yasuhiro Ikabata, Junji Seino, Takeshi Yoshikawa, Ryo Fujisawa, Hiromi Nakai
2. 発表標題 Development of the machine-learned correlation model: systematic assessment on chemical properties
3. 学会等名 日本化学会第100春季年会
4. 発表年 2020年

1. 発表者名 藤澤遼, 五十幡康弘, 清野淳司, 吉川武司, 中井浩巳
2. 発表標題 凍結内殻近似に基づく機械学習型電子相関モデルの開発
3. 学会等名 日本化学会第100春季年会
4. 発表年 2020年

1. 発表者名 五十幡康弘, 棚嶋拓朗, 清野淳司, 吉川武司, 藤澤遼, 中井浩巳
2. 発表標題 機械学習型電子相関モデルの開発と数値検証
3. 学会等名 第42回ケモインフォマティクス討論会
4. 発表年 2019年

1. 発表者名 五十幡康弘, 吉川武司, 中井浩巳
2. 発表標題 相対論的量子化学計算プログラムRAQETの公開
3. 学会等名 日本コンピュータ化学会2019年春季年会
4. 発表年 2019年

1. 発表者名 五十幡康弘, 棚嶋拓朗, 清野淳司, 吉川武司, 中井浩巳
2. 発表標題 相関エネルギー密度の完全基底極限に基づく機械学習型電子相関モデルの開発
3. 学会等名 第22回理論化学討論会
4. 発表年 2019年

1. 発表者名 櫛嶋 拓朗, 五十幡 康弘, 清野 淳司, 影山 椋, 吉川 武司, 中井 浩巳
2. 発表標題 機械学習を用いたpost-Hartree-Fock電子相関モデルの開発
3. 学会等名 日本化学会第99春季年会
4. 発表年 2019年

1. 発表者名 櫛嶋 拓朗, 五十幡 康弘, 清野 淳司, 影山 椋, 藤波 美起登, 中井 浩巳
2. 発表標題 機械学習を用いた交換相関汎関数の開発
3. 学会等名 第12回分子科学討論会
4. 発表年 2018年

1. 発表者名 五十幡 康弘, 櫛嶋 拓朗, 清野 淳司, 影山 椋, 藤波 美起登, 中井 浩巳
2. 発表標題 Post-Hartree-Fock相関エネルギー密度の機械学習による相関エネルギー計算手法の開発
3. 学会等名 ポスト「京」第5回公開シンポジウム
4. 発表年 2018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
---------	---------