

分子軌道エネルギーと機械学習による分子物性の予測

寺前 裕之^{a*}, 松尾 哲秀^a, 庭月野 一真^a, 井上 竜太^a, 野口 晋治^a 玄美燕^b, 山下 司^b, 高山 淳^b,
岡崎 真理^b, 坂本 武史^b

^a城西大学理学部化学科, 〒350-0295 坂戸市けやき台1-1

^b城西大学薬学部薬学科, 〒350-0295 坂戸市けやき台1-1

*e-mail: teramae@gmail.com

(Received: July 26, 2020; Accepted for publication: September 17, 2020; Online publication: November 19, 2020)

The values of the internuclear distances and the dipole moments of 14 small molecules have been estimated by machine learning with only molecular orbital energies as the explanatory variables. We use four regression methods, partial least square (PLS), random forest (RF), Radial Basis Function Kernel Regularized Least Squares (krlsRadial), and Bayesian Regularized Neural Networks (BRNN) and we report only BRNN results for the internuclear distances, and PLS results for the dipole moments. The coefficients of determination for the internuclear distances and the dipole moments are 0.9318 and 0.7265, respectively. It has been proved that the internuclear distances and the dipole moments can be predicted by the molecular orbital energies only.

Keywords: Dipole moments, Equilibrium geometries, Eigenvalues of molecular orbital, Machine learning, Bayesian regularized neural networks

1 はじめに

寺前らは日本で認可・販売されているベンゾジアゼピンおよびチエノジアゼピン系の抗不安薬について, *ab initio*分子軌道法によるエネルギーとエネルギー勾配を用いた高次元アルゴリズムによる分子動力学計算を行い, 薬物の抗不安性と抗痙攣性の強さと分子軌道のエネルギー値に相関があることを初めて見いだした [1].

HF/3-21G 計算によりコンフォメーション解析ならびに構造最適化を行い, 得られた構造における HF/6-31G (d, p)分子軌道のエネルギー固有値と抗不安性や抗痙攣性の強度を比較したところ, いずれも next-HOMO の軌道エネルギー固有値との間に強い相関関係を見いだした.

近年, 寺前ら [2]はフェルラ酸(FA)の抗酸化作用 [3, 4]を同様な手法を用いることで, 説明できることを示した. DPPHフリーラジカル消去濃度(IC₅₀)とフェノール性水酸基から水素を取ったFAラジカルの電子状態と IC₅₀との関連性が考察された. ただし抗不安薬の場合と異なり, ただ一つの軌道エネルギー固有値では強い相関関係を得ることはできなかったため, ラジカル種とニュートラル種の各軌道エネルギー値と IC₅₀との関係を random forest モデルにより機械学習を試みたところ, α

SOMO-1が6%, α SOMOが24%, α LUMOが45%の寄与があると計算された [2].

このように抗不安性や抗酸化性などのように複雑な過程の相関関係を簡単に得ることができるので, より簡単な分子物性を分子軌道計算と機械学習により, 相関関係を得ることが可能ではないかと考えた. 本研究では, 2原子分子 ClF, HCl, HF, CS, LiF, LiCl, NaF, NaCl, N₂, NP, SiO, F₂, P₂, Cl₂ の14種類の平衡核間距離および異核2原子分子, ClF, HCl, HF, CS, LiF, LiCl, NaF, NaCl と PH₃, NH₃, HCP, H₂S, H₂O, H₂CO の14分子の双極子モーメントを分子軌道エネルギー値のみを説明変数として機械学習を試みた.

2 計算方法

分子軌道計算プログラム Gaussian16 を使用し [5], 6-31G (d, p)基底を用いて構造最適化を行った. 構造最適化を行った構造による分子軌道固有値を説明変数として用いた. 機械学習は R 言語により多くの回帰法が実装されている caret パッケージ [6]を用いた. 核間距離と双極子モーメントについての文献値は NIST のデータベースのものを用いた [7].

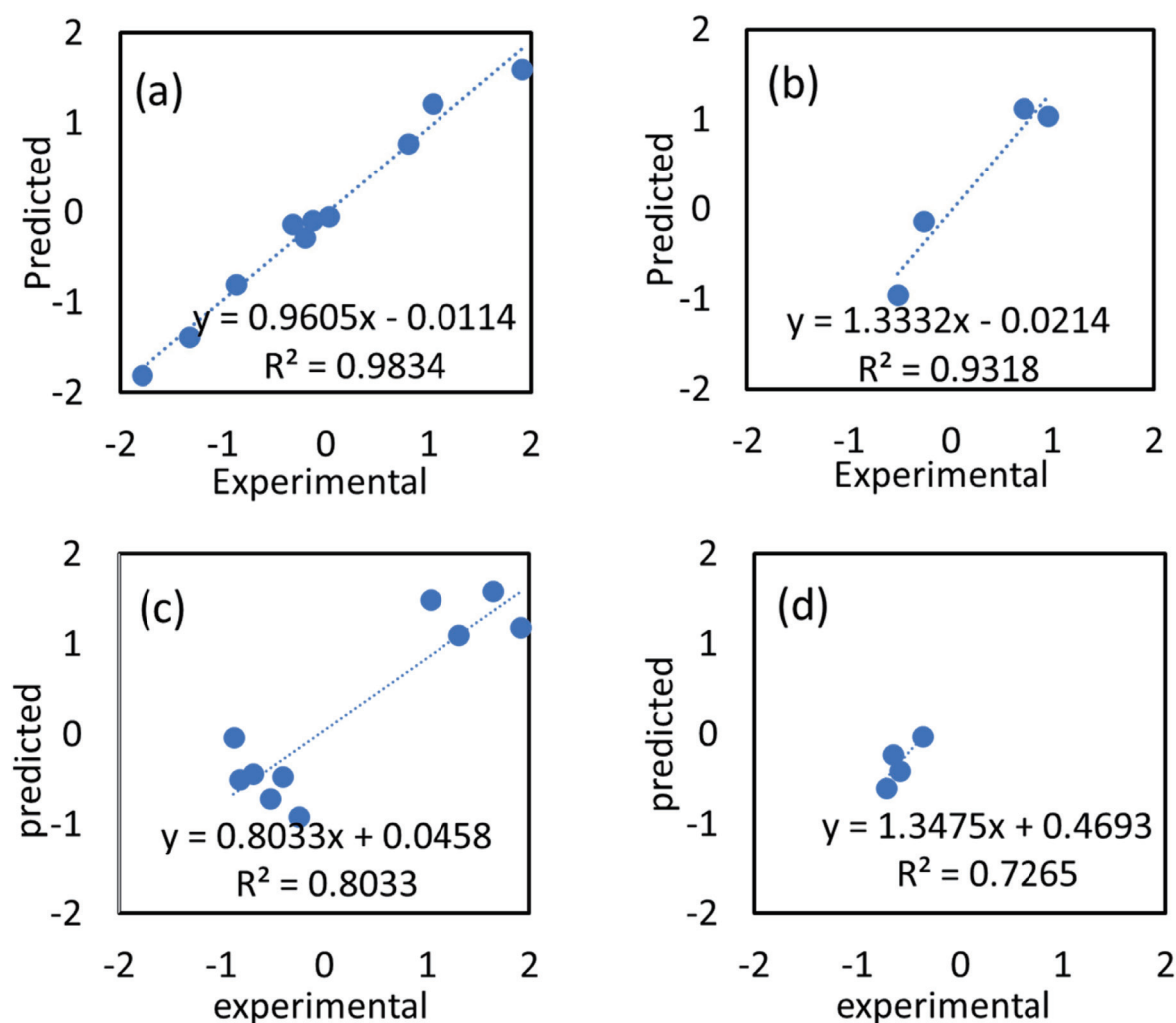


Fig.1 Plots of Training and Predicted Values of (a), (b) Internuclear Distances and (c), (d) Dipole Moments. Regression Methods are (a), (b) BRNN and (c), (d) PLS, respectively,

3 結果と考察

機械学習における回帰法としては、partial least square (PLS), random forest (RF), Radial Basis Function Kernel Regularized Least Squares (krlsRadial) と Bayesian Regularized Neural Networks (BRNN) の4種類について試みた。

基底関数の関係から、同一レベルで計算できる分子数が14種類と限られているが、最初に10+4の2つのグループに乱数を使用して分割し、前者で教師付の機械学習を行い、その結果を用いて後者で予測を行い、どの程度の精度が得られているかを確認した。分割はいくつかのパターンを試行して大きな差が出ないことを確認した。また過学習を避けるために教師付学習においては Validation を行っている。

説明変数としての軌道エネルギーの数は核間距離の場

合で被占軌道ではHOMOから下の5軌道とLUMOから上の5軌道、合計10軌道とした。これは計算された分子の中で最も電子数の少ないHF分子の電子数から決定した。双極子モーメントの方はHOMOから下の3軌道、LUMOから上の3軌道、合計6軌道とした。こちらについては特に理由は無く適当に決定した。

ここでは、最も良く文献値に合致していた回帰法、核間距離ではbrnn、双極子モーメントではplsを用いた結果のみを報告する。ただし他の回帰法を用いた場合も概ね同様な結果が得られている。

Figure 1に機械学習の結果を示した。Figure 1 (a)に核間距離の教師付機械学習、Figure 1 (b)には、核間距離の予測値、Figure 1 (c)に双極子モーメントの教師付機械学習、Figure 1 (d)には双極子モーメントの予測値を示し

た。教師付機械学習の結果は核間距離は、決定係数 R^2 が0.9834、双極子モーメントの場合で0.8033といずれも強い相関関係を示している。特に双極子モーメントの場合はさらに複雑な回帰法、例えばkrlsRadial, により、文献値と完全に合致させることも可能であるが、過学習により予測値は却って合致しなくなってしまうことには注意を要する。

予測値の決定係数では、核間距離の場合で、0.9318、双極子モーメントで0.7265となっている。核間距離はほぼ完全に予測可能、双極子モーメントもかなりの確率で予測が可能であることを示している。

4 結論

機械学習を用いていくつかの小さな分子の核間距離と双極子モーメントの値を分子軌道のエネルギー値を説明変数として予測することを試みた。

4種類の回帰法を使用した。いずれの方法でもある程度の相関が得られ、分子軌道計算と機械学習の組み合わせは強力なツールであり、様々な分子物性の予測が期待できるであろうことがわかった。ただし、本研究の核間距離や双極子モーメントはあくまでテスト目的であり、実用性はないため、エントロピー、モル熱容量、融点、沸点などの、分子物性への適用を今後行っていく予定にしている。

参考文献

- [1] H. Teramae, K. Ohtawara, T. Ishimoto, U. Nagashima, *Bull. Chem. Soc. Jpn.*, **81**, 1094 (2008). DOI:10.1246/bcsj.81.1094
- [2] H. Teramae, M. Xuan, T. Yamashita, J. Takayama, M. Okazaki, T. Sakamoto, *J. Comput. Chem. Jpn.*, **18**, 211 (2019). DOI:10.2477/jccj.2019-0034
- [3] K. H. Chang, M. L. Cheng, M. C. Chiang, C. M. Chen, *Clin. Chim. Acta*, **485**, 79 (2018). DOI:10.1016/j.cca.2018.06.031, PMID:29940147
- [4] M. Xuan, M. Okazaki, N. Iwata, S. Asano, S. Kamiuchi, H. Matsuzaki, T. Sakamoto, Y. Miyano, H. Iizuka, Y. Hibino, *Evid. Based Complement. Alternat. Med.*, **2015**, 1 (2015). DOI:10.1155/2015/865986, PMID:25945116
- [5] Gaussian 16, Revision B.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.
- [6] <https://cran.r-project.org/web/packages/caret/caret.pdf>
- [7] NIST Computational Chemistry Comparison and Benchmark Database NIST Standard Reference Database Number 101 Release 20, August 2019, Editor: Russell D. Johnson III, <http://cccbdb.nist.gov/> DOI:10.18434/T47C7Z