

機械学習による完全基底関数極限における電子相関エネルギーの推定

(早大理工研¹、早大先進理工²、JST-CREST³、京大 ESICB⁴)○清野淳司¹、大越昌樹^{1,4}、中井浩巳^{1,2,3,4}Estimation of complete basis set limit in electron correlation energy
utilizing machine learning(RISE, Waseda Univ.¹, Advanced Science and Engineering, Waseda Univ.², JST-CREST³, ESICB, Kyoto Univ.⁴)○Junji Seino¹, Masaki Okoshi^{1,4}, and Hiromi Nakai^{1,2,3,4}

【緒言】 化学的精度 (<2.5 kJ/mol) で量子化学計算を行うために、完全基底関数 (CBS) 極限における高精度な電子相関エネルギーを算出することは重要である。これまで電子相関エネルギーが、Dunning の correlation-consistent 基底関数の基数 X に対して、 X^{-3} で CBS 極限に収束することが示され^[1]、さらにこの収束性を拡張した外挿手法が提案されてきた^[2-5]。これにより、低コストの DZ ($X=2$) および TZ ($X=3$) 基底関数を用いて、CBS 極限における電子相関エネルギーの高精度な推定が可能になった。本研究では機械学習を用いることで、従来の外挿法の等価性を証明し、さらにその関数形を拡張することで、より高精度なフィッティング手法を開発した。また、いくつかの電子相関理論と基底関数の組による結果を機械学習により組み合わせた複合法も提案し、その有効性を検証した。

【従来の外挿法の等価性の証明】 これまで提案されてきた CBS 極限における電子相関エネルギーの一般的な関数形は以下の通りである。

$$E^{\text{CBS}}[X, X+n] = \frac{(X+n+\alpha)^\beta E[X+n] - (X+\gamma)^\beta E[X]}{(X+n+\alpha)^\beta - (X+\gamma)^\beta} \quad (1)$$

ここで、 (α, β, γ) はそれぞれの電子相関理論/外挿法に依存する外挿のパラメータである。例えば Helgaker らの手法 (HKKN)^[1] ではすべての電子相関理論で (0.0, 3.0, 0.0) で固定されており、Huh と Lee の手法 (HL)^[2] では、MP2 法で (1.0, 3.0, 1.0)、CCSD 法で (0.5, 3.0, 0.5) と決定されている。しかし、これまでの手法は独自のテストセットでパラメータ決定および数値検証がなされており、どの手法が最も高精度か検証されていない。そこで本研究では、MP2 法と CCSD 法に対して Gaussian-3X セットの 223 分子を、CCSD(T)法に対して Gaussian-2 セットの 148 分子を用いて、それぞれの外挿法のパラメータを最小二乗法により再決定し、数値検証を行った。Table 1 に HKKN、HL、Bakowies の手法 (Bkw)^[3]、Varandas と Pansini の手法 (VP)^[4]、我々の以前の手法 (OAN(C))^[5] について、これまで (old) と再決定された (rev) パラメータを用いた、 $n=1$ 、 $X=2$ における平均絶対偏差 (MAD) を示す。参照値として $n=$

Table 1. MADs from HKKN ($n=1$, $X=3$) (kJ/mol).

	HKKN	HL		Bkw		VP		OAN(C)	
		old	rev	old	rev	old	rev	old	rev
MP2	73.8	42.1	7.7	21.8	7.7	14.6	7.7	7.7	7.7
CCSD	49.9	11.7	11.4	11.6	11.4	12.0	11.4	11.1	11.4
CCSD(T)	41.8	10.1	10.0	10.1	10.0	10.8	10.0	10.8	10.0

1, $X=3$ における HKKN の結果を用いた。この結果、old では手法によ

って MAD が異なるが、同じセットでパラメータを決定した rev では、すべての手法で同じ結果を与えることが確認された。これは、HKKN 以外の手法が、次式の線形外挿で表されるためである。

$$E_{\text{Linear1}}^{\text{CBS}}[X, X+n] = aE[X+n] + (1-a)E[X] \quad (a > 1) \quad (2)$$

【外挿法の拡張】式(2)の関数形を変えることで、より一般的な関数形を用いることができる。

$$E_{\text{Linear2}}^{\text{CBS}}[X, X+n] = aE[X+n] + bE[X] \quad (a+b \neq 1) \quad (3)$$

$$E_{\text{Quad}}^{\text{CBS}}[X, X+n] = aE[X+n] + bE[X] + cE[X+n]^2 + dE[X]^2 + eE[X+n]E[X] \quad (4)$$

$$E_{\text{NN}}^{\text{CBS}}[X, X+n] = \text{Neural network}(E[X+n], E[X]) \quad (5)$$

ここで、式(3)では式(2)の係数の制約を緩和しており、式(4)ではエネルギーの2乗の項を加えることで非線形性を表現する。式(5)は2点のエネルギーを入力とした、機械学習の一つであるニューラルネットワーク (NN) である。NN は一般的に任意の関数形が表現できる。Table 2 に、それぞれの手法の $n=1$ 、 $X=2$ における MAD を示す。

Table 2. MADs from HKKN ($n=1$, $X=3$) (kJ/mol).

	Linear1	Linear2	Quad	NN
MP2	7.7	5.2	4.9	4.8
CCSD	11.4	7.2	5.9	5.6
CCSD(T)	10.0	6.7	5.1	4.8

この結果、式(3)の係数制約の緩和により、さらに式(4)と式(5)の非線形性の導入により、改善されることが確認された。Quad や NN では、MAD は 4-6 kJ/mol である。

【複合法】次に、いくつかの基底関数における MP2 法、CCSD 法と CCSD(T)法のエネルギーを入力とした、NN による CCSD(T)/CBS の推定法を提案する。

$$E_{\text{CCSD(T)}}^{\text{CBS, NN}} = \text{Neural network}(E_{\text{MP2}}[i](i=2, \dots, l), E_{\text{CCSD}}[j](j=2, \dots, m), E_{\text{CCSD(T)}}[k](k=2, \dots, n)) \quad (6)$$

Figure 1 に Gaussian-2 セットにおける、式(6)の $l=4$, $m=3$, $n=2$ での誤差のヒストグラムと MAD を示す。式(5)の結果も合わせて載せる。この手法は CCSD(T)法の基底関数が DZ ($X=2$) であるため、特に大きな系において、より効率的に CCSD(T)/CBS を推定できる。この結果、式(5)の手法では全体の 13% の分子の誤差が 10 kJ/mol 以上あるが、式(6)ではすべての分子で 6 kJ/mol 以下であり、高精度な結果を与えることが示された。

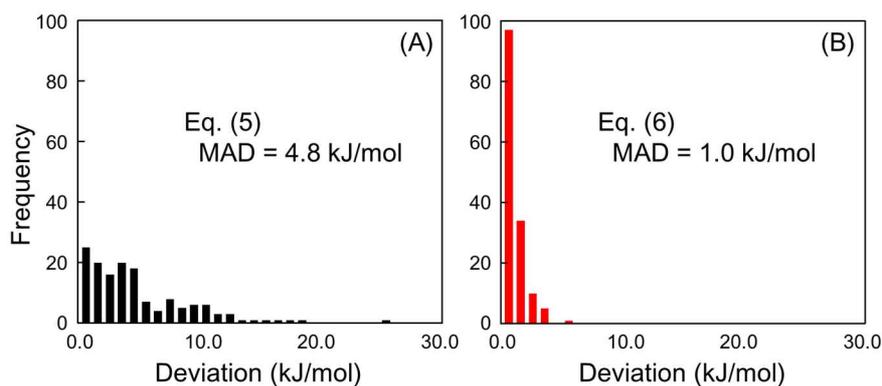


Figure 1. Histograms of deviations (kJ/mol) of schemes in (A) Eq. (5) and (B) Eq. (6) ($l=4$, $m=3$, $n=2$) using neural network from CCSD(T)/HKKN ($n=1$, $X=3$).

【参考文献】 [1] T. Helgaker, W. Klopper, H. Koch, and J. Noga, *J. Chem. Phys.* **106**, 9639 (1997). [2] S. B. Huh and J. S. Lee, *J. Chem. Phys.* **118**, 3035 (2003). [3] D. Bakowies, *J. Chem. Phys.* **127**, 084105 (2007). [4] A. J. C. Varandas and F. N. N. Pansini, *J. Chem. Phys.* **141**, 224113 (2014). [5] M. Okoshi, T. Atsumi, and H. Nakai, *J. Comput. Chem.* **36**, 1075 (2015).