計算工学講演会論文集　Vol.28 (2023年5月)　計算工学会

DPDシミュレーション用パラメータの非経験的算定の機械学習による効率化

Improving Efficiency of Non-Empirical Calculation of Parameters for DPD Simulation by Machine Learning

土居英男[1)]，松岡壮太[2)]，奥脇弘次[3)]，畑田崚[2)]，
南聡次朗[4)]，栖原涼輔[4)]，望月祐志[5,6)]

Hideo Doi, Sota Matsuoka, Koji Okuwaki, Ryo Hatada,
Soujiro Minami, Ryosuke Suhara and Yuji Mochizuki

1) 博(理) 立教大学理学部化学科 助教（〒171-8501 東京都豊島区 西池袋3丁目34-1, E-mail: hideo-doi@rikkyo.ac.jp)
2) 学(理) 立教大学理学研究科化学専攻 大学院生（〒171-8501 東京都豊島区 西池袋3丁目34-1)
3) 博(理) 立教大学理学部化学科 研究員（〒171-8501 東京都豊島区 西池袋3丁目34-1)
4) 学(理) 立教大学理学部化学科 学部生（〒171-8501 東京都豊島区 西池袋3丁目34-1)
5) 理博 立教大学理学部化学科 教授（〒171-8501 東京都豊島区 西池袋3丁目34-1, E-mail: fullmoon@rikkyo.ac.jp)
6) 理博 東京大学生産技術研究所 リサーチフェロー (〒153-8503 東京都目黒区駒場4-6-1)

In recent years, the demand for advanced materials development has increased in the field of materials science. Dissipative Particle Dynamics (DPD) simulations are used to study materials. However, these simulations require parameters that describe the interactions between DPD particles. Therefore, we sought to optimize the calculation of χ parameters for DPD simulations by machine learning.

Key Words : *DPD, FMO, Machine Learning, Interaction energy, Chi parameter*

1. はじめに

近年，マテリアルサイエンスや創薬の分野で高度な機能を有した材料開発の要求が高まっている．メゾスケールでの構造や振る舞いの予測・制御が重要であるためである．通常，メゾ構造の予測には散逸粒子動力学（DPD）や粗視化分子動力学のシミュレーションを用いられるが，これらのシミュレーションの実行には分子や粒子同士の相互作用を表現するパラメータが必要である．これらのパラメータは実験データを元にフィッティングする[1, 2]か古典的な分子力学（MM）にもとづいて算定[3]されることがほとんどであった．実験データは時間的、金銭的コストが高く、MMに基づく算定は力場と現実の齟齬が大きな問題となる。そこで，我々は，フラグメント分子軌道(FMO)計算に基づいて非経験的[4, 5]にDPDシミュレーションのためのパラメータを算定するソフトウェアFCEWS(FMO-based Chi parameter Evaluation Workflow System)[6, 7]を開発した．

FCEWSでは，分子をある程度の分子セグメントに分割し，その分子セグメント毎に数千構造の非経験的分子軌道法による相互作用エネルギーの計算を行う．この相互作用計算は分子セグメントのペア毎に行う必要があり，非経験的分子軌道法を用いて相互作用計算を行っているため，多大な計算リソースと時間が必要である．

近年，機械学習の進歩により，相互作用計算を高速化するための新しい方法が開発されており，製薬で重要であるタンパク質とのバインディングエネルギーや毒性など広範な目的のために使用されている[8]. Chemielaらは，非経験的分子軌道法の結果などから機械学習を非経験的分子軌道法のシミュレーションの結果から分子動力学シミュレーションのためのポテンシャルを自動で生成し評価した[9]．また，Alibakhshiらは機械学習によって溶媒和自由エネルギーの予測精度を向上させた[10]．機械学習で部分電荷を予測する取り組みも行われている[11–13]．

そこで，我々は機械学習を用いてDPDシミュレーションのためのχパラメータの算定の効率化を試みた．

2. FCEWSにおけるχパラメータの算定

χパラメータはFlory-Hugginsの格子理論で定義される粒子間の相互作用パラメータである．FCEWSでは，ある分子セグメント1と分子セグメント2の間のχパラメータをどのように計算しているのかを手短に説明する．詳しくは文献[7]を参照されたい．

$$\chi = \frac{(\boldsymbol{Z_{12}S_{12}E_{12}} + \boldsymbol{Z_{21}S_{21}E_{21}}) - (\boldsymbol{Z_{11}S_{11}E_{11}} + \boldsymbol{Z_{22}S_{22}E_{22}})}{\boldsymbol{RT}} \quad (1)$$

© 一般社団法人日本計算工学会

ここで，配位数$\boldsymbol{Z_{12}}$は分子セグメント1を中心に配置した時，分子セグメント2を最大でいくつ配置できるかの変数である．スケーリングファクター$\boldsymbol{S_{12}}$は系の異方性を表現するパラメータである．エネルギー$\boldsymbol{E_{12}}$は分子セグメント1と分子セグメント2の平均的な相互作用エネルギーである．$\boldsymbol{E_{21}}, \boldsymbol{E_{11}}, \boldsymbol{E_{22}}$も同様である．エネルギー$\boldsymbol{E_{12}}$は式(2)で計算する．

$$\boldsymbol{E_{12}} = \frac{\boldsymbol{\Sigma E_i exp(-E_i/RT)}}{\boldsymbol{\Sigma exp(-E_i/RT)}} \tag{2}$$

エネルギーE_iはi番目の分子セグメント1と分子セグメント2の二量体の相互作用エネルギーで，定数Rは気体定数であり，変数Tは温度である．通常，この時計算される二量体の個数は2000個である．スケーリングファクター$\boldsymbol{S_{12}}$は以下のように計算する．ある分子セグメント1を原点に配置し，その周囲に標本点と呼ばれる点を6個配置する．座標は(1,0,0),(-1,0,0),(0,1,0),(0,-1,0),(0,0,1),(0,0,-1)をランダムに回転させた点である．分子セグメント1の周囲に分子セグメント2が配置され，あるi番目の二量体の構造があり，分子セグメント2の中心から最近接の標本点を$j(j = 1, \ldots, 6)$とすると，各標本点の重みWは式(3)で計算できる．

$$\boldsymbol{W_j = \Sigma \mathrm{exp}\,(-E_i/RT)} \tag{3}$$

重みW_jは，空間を6分割し，ある空間に存在する分子セグメント1と2の間に存在する相互作用エネルギーのボルツマン因子の総和となっている．スケーリングファクター$\boldsymbol{S_{12}}$は，式(4)で計算する．

$$\boldsymbol{S_{12}} = \sum_{j=1}^{6} \frac{\boldsymbol{W_j}}{\boldsymbol{Max(W)}} \tag{4}$$

このスケーリングファクターは，セグメント1が等方的に相互した場合は1に近くなり，特定の結合サイトと強く相互作用する場合は小さくなるような変数である．

上記で示したように，ある分子セグメントの組み合わせの二量体の相互作用エネルギーが計算，もしくは予測しさえできれば，χパラメータは計算が可能である．

3. 機械学習手法

本研究では，Scikit-learn (version 0.24.2) [14]に実装されているRandom Forest [15]を用いて回帰分析を行った．なお，今回はハイパーパラメータの最適化は行わず，max_depth (決定木のノード深さ)を80，n_estimators (推定器の数)を50に，その他をデフォルトで分析を行った．

4. データセット

本研究では，ヘキサン-ニトロベンゼンの系を予測対象に選び，ヘキサン-ヘキサン(H-H)，ヘキサン-ニトロベンゼン(H-N)，ニトロベンゼン-ニトロベンゼン(N-N)の二量体の相互作用エネルギーの予測を試みた．

教師データとして，ABINIT-MP[16, 17]を使用し各分子の二量体をFMO-MP2/6-31G†レベルで計算し，モノマーのエネルギーとの差を相互作用エネルギーとした．各分子のモノマーの構造は，GAUSSIAN16[18]用いてB97D/6-31G(d)レベルで最適化を行った．1つの分子の組み合わせ毎に2000構造の二量体構造を生成し，計6000構造の相互作用エネルギーを使用した．

記述子として，以下の5種類を用意した．1: クーロン相互作用D_qを式(5)で計算した．モノマーのRESP電荷[19–21]を使用し，静電相互作用を計算した．RESP電荷はABINIT-MPを使用して算定した．

$$\boldsymbol{D_q} = \sum \frac{\boldsymbol{q_i q_j}}{\boldsymbol{4\pi\epsilon_0 r_{ij}}} \tag{5}$$

変数q_i, q_jは各分子における原子i, jの電荷であり，距離r_{ij}は原子i, j間の距離であり，定数ϵ_0は真空中の誘電率である．2: レナード-ジョーンズポテンシャルの記述子D_{LJ}として，式(6)で計算を行った．力場として，GAFF [22]，Ghemical [23]，UFF [24]を使用し，計算は，OpenBabel [25]で行った．

$$\boldsymbol{D_{LJ}} = \sum \boldsymbol{4\epsilon}\left[\left(\frac{\boldsymbol{\sigma}}{\boldsymbol{r_{ij}}}\right)^{\boldsymbol{12}} - \left(\frac{\boldsymbol{\sigma}}{\boldsymbol{r_{ij}}}\right)^{\boldsymbol{6}}\right] \tag{6}$$

定数σ, ϵともに，各力場での固有のパラメータである．3: 原子間距離の記述子D_dとして，各原子の種類毎に式(7)で計算を行った．

$$\boldsymbol{D_d} = \sum \left(\frac{\boldsymbol{1}}{\boldsymbol{r_{ij}}}\right)^{\boldsymbol{n}} \tag{7}$$

冪数nとして，{3,4,5,6,8,10,12}を使用し，原子種として，{H,C,O,N}を，原子種の組み合わせとして{H-H,H-C,H-N,H-O,…,O-O}を使用した．4:各二量体構造における，各分子の幾何中心間の距離を記述子D_cとして使用した．5: 原子種の組み合わせごとの最近接原子間距離の記述子としてD_nを使用した．

5. 結果と考察

χパラメータの算出に必要なのは相互作用エネルギーのみである．そこで本研究では，相互作用エネルギーの予測に必要なFMO計算の回数と，その時の予測精度を調査した．事前に2000回のFMO計算を行い，そのうち学習に

© 一般社団法人日本計算工学会

使用するデータ数を250, 500, 1000, 1500と変化させ，二量体の相互作用エネルギー予測を行った．予測の精度評価は5-fold cross validationでのR^2を使用し，また学習に使用しなかったデータもQ^2で評価した．結果を表-1に示した．

全体的に，データ数が多くなるにつれて予測精度が高くなっていた．また，予測精度は分子の組み合わせの影響を大きく受け，非極性分子であるHexaneペア(H-H)の予測精度はデータ個数が250程度であってもR^2が0.94程度と高い精度だったが，極性分子であるNitrobenzeneを含むペア(H-N, N-N)での予測精度はどちらも低く，データ個数を1500まで上げても0.90程度であった．データ個数を250から1500にしたところ，どの分子ペアでもMAEは概ね2/3程度の予測精度向上が見られた．学習データでのR^2と非学習データのQ^2を比較した結果，両者はほとんど同じであり過学習などは起こっていないと思われた．図-1に、N-Nでの1500個のデータを使用しての学習結果におけるテストデータの散布図を示す。

表-1 各分子ペアでの予測精度に対する学習データの数の影響 (a) 学習データでのR^2の値

(b) 非学習データでのQ^2の値

(a)

	250	500	1000	1500
H-H	0.941	0.963	0.971	0.976
H-N	0.834	0.864	0.903	0.907
N-N	0.841	0.873	0.913	0.917

(b)

	250	500	1000	1500
H-H	0.953	0.959	0.965	0.968
H-N	0.857	0.870	0.881	0.889
N-N	0.859	0.867	0.883	0.893

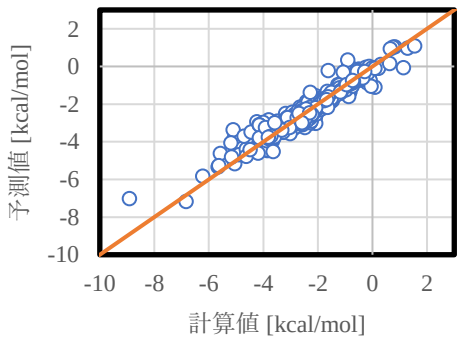

図-1 N-N における計算値と予測値の散布図

次に，上記で予測した個々の相互作用エネルギーを使用して，平均相互作用エネルギーE_{ij}を計算し，学習数別で比較した．なお，平均相互作用エネルギーを計算する時の温度として300 Kを設定した．結果を表-2に示す．

表-2 300 K での平均相互作用エネルギーE_{ij}の予測精度に対する学習データの数の影響

	250	500	1000	1500	2000
H-H	-1.48	-1.48	-1.48	-1.49	-1.50
H-N	-2.85	-3.22	-3.40	-3.48	-3.73
N-N	-4.74	-5.88	-6.98	-7.85	-8.06

全ての分子の組み合わせで，予測値は2000構造で計算された値よりも高い値が算出された．また，学習データを増やすと誤差は減少しった．H-Hでは，250構造を使用しての予測値は-1.48と計算値の-1.50とほぼ同じだったが，H-N, N-Nでは，予測値と計算値の差は非常に大きかった．平均相互作用エネルギーを計算する時のボルツマン因子は，エネルギーが低いほど重みが大きくなり，数値に与える影響が大きいため，相互作用エネルギーが低い場合，平均エネルギーに対して大きな影響を与える．しかし，そのような構造は数が少ない．特に，H-N, N-Nの組み合わせのように，極性分子が存在し，分子が特定の配置の場合，安定な相互作用エネルギーが計算されており，そのため，教師データが不足しており，機械学習による予測が難しく，このような結果になったと考えられる．

次に相互作用エネルギーの予測値のスケーリングファクターS_{ij}への影響を調べた．結果を表-3に示す．こちらも，平均相互作用エネルギーの場合と同様で，スケーリングファクターの誤差は非常に大きかった．

表-3 300 K でのスケーリングファクターS_{ij}の予測精度に対する学習データの数の影響

	250	500	1000	1500	2000
H-H	0.79	0.78	0.78	0.78	0.78
H-N	0.72	0.62	0.62	0.59	0.54
N-H	0.54	0.49	0.50	0.48	0.44
N-N	0.50	0.45	0.40	0.34	0.33

表-4 300 K で算定されたχパラメータの予測精度に対する学習データの数の影響

	250	500	1000	1500	2000
χ	1.1	3.3	2.5	1.3	2.0

次に相互作用エネルギーE_{ij}とスケーリングファクターS_{ij}を使用してχパラメータを計算し，機械学習の予測値と実際の計算値との比較を行った．結果を表-4に示す．なお，このヘキサン－ニトロベンゼンの系は，293 K付近でχパラメータが2.0になることが実験的にわかっており[1]，300 Kで約2.0になることがわかっている．予測値を使用した場合，データ数に関係なく誤差が大きくなった．全体の内の75%の構造である1500個の構造の相互作用エネルギーをFMO計算していたとしても，予測したχパラメータにおける誤差が非常に大きく，予測値なしで計算した2.0とは大きく異なっていた．これは，DPDシミュレーションに

© 一般社団法人日本計算工学会

使用するためのχパラメータとして不適切な水準であった．全体の内，25%である500個の構造の相互作用エネルギーに予測値を使用するだけでχパラメータに大きな影響を与えるということは，素朴なアプローチだけでは，機械学習の予測値を使用することができないことを意味し，何か他のアプローチを取る必要があることが明らかになった．一方で，このような機械学習を利用し，χパラメータの予測を行うことが可能であることを示せた意義は大きく，今後の進歩が望まれる結果となった．

6. 謝辞

本研究は，立教SFRから支援を受けました．また，本研究開発にあたり，JSOL(株)の小沢拓氏との情報交換・議論が有益であることを記します．

参考文献

[1] Drozd-Rzoska, A. et al.: Liquid–liquid phase equilibria in nitrobenzene–hexane critical mixture under negative pressure. *Phys. Chem. Chem. Phys.*, Vol.6, pp. 2291–2294, 2004.

[2] Polymer Property Predictor and Database (https://pppdb.uchicago.edu/chi)

[3] Fan, C.F. et al.: Application of molecular simulation to derive phase diagrams of binary mixtures. *Macromolecules*, Vol.25, pp. 3667–3676, 1992.

[4] "The Fragment Molecular Orbial Method: Practical Applications to Large Molecular Systems", ed. D.G. Fedorov and K. Kitaura, CRC Press (2009).

[5] "Recent Advances of the Fragment Molecular Orbital Method - Enhanced Performance and Applicability", ed. Y. Mochizuki, S. Tanaka and K. Fukuzawa, Springer (2021).

[6] Okuwaki, K. et al.: An automated framework to evaluate effective interactionparameters for dissipative particle dynamics simulations basedon the fragment molecular orbital (FMO) method. *J. Comput. Chem. Jpn.*, Vol.17, pp. 102–109, 2018.

[7] 奥脇弘次，土居英男，望月裕志 :フラグメント分子軌道(FMO)法を用いた散逸粒子動力学シミュレーションのための有効相互作用パラメータ算出の自動化フレームワーク. *J. Comput. Chem. Jpn.*, Vol.17, pp. 102–109, 2018.

[8] Wu, Z. et al.: MoleculeNet: a benchmark for molecular machine learning. *Chem. Sci.*, Vol.9, pp. 513–530, 2018.

[9] Chmiela, S. et al.: Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.*, Vol.9, pp. 3887, 2018.

[10] Alibakhshi, A., Hartke, B.: Improved prediction of solvation free energies by machine-learning polarizable continuum solvation model. *Nat. Commun.*, Vol.12, pp. 3584, 2021.

[11] Bleiziffer, P. et al.: Machine learning of partial charges derived from high-quality quantum-mechanical calculations. *J. Chem. Inf. Model.*, Vol.58, pp. 579–590, 2018.

[12] Kato, K. et al.: High-precision atomic charge prediction for protein systems using fragment molecular orbital calculation and machine learning. *J. Chem. Inf. Model.*, Vol.60, pp. 3361–3368, 2020

[13] Doi, H. et al.: Screening toward the Development of Fingerprints of Atomic Environments Using Bond-Orientational Order Parameters. *ACS Omega*, Vol.7, pp. 4606–4613, 2022.

[14] Pedregosa, F. et al.: Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.*, Vol.12, pp.2825–2830, 2011.

[15] Breiman, L.: Random Forests. *Mach. Learn.*, Vol.45, pp. 5–32, 2001.

[16] Tanaka, S. et al.: Electron-correlated fragment-molecular-orbital calculations for biomolecular and nano systems. *Phys. Chem. Chem. Phys.*, Vol.16, pp. 10310–10344, 2014.

[17] Y. Mochizuki et al., "The ABINIT-MP Program" in Ref. [5], pp. 53-67.

[18] Frisch, M.J. et al.: *Gaussian16 Revision B.01*. 2016.

[19] Cornell, W.D. et al.: Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.*, Vol.115, pp. 9620–9631, 1993.

[20] Okiyama, Y. et al.: Application of the fragment molecular orbital method for determination of atomic charges on polypeptides. *Chem. Phys. Lett.*, Vol.449, pp. 329–335, 2007.

[21] Okiyama, Y. et al.: Application of the fragment molecular orbital method for determination of atomic charges on polypeptides. II. Towards an improvement of force fields used for classical molecular dynamics simulations. *Chem. Phys. Lett.*, Vol.467, pp. 417–423, 2009.

[22] Wang, J. et al.: Development and testing of a general amber force field. *J. Comput. Chem.*, Vol.25, pp. 1157–1174, 2004.

[23] Hassinen, T., Peräkylä, M.: New energy terms for reduced protein models implemented in an off-lattice force field. *J. Comput. Chem.*, Vol.22, pp. 1229–1242, 2001.

[24] Rappe, A.K. et al.: UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.*, Vol.114, pp. 10024–10035, 1992.

[25] O'Boyle, N.M. et al.: Open Babel: An open chemical toolbox. *J. Cheminform.*, Vol.3, pp. 33, 2011.

© 一般社団法人日本計算工学会