

## 課題名 (タイトル) :

第一原理大規模データ及び機械学習に依る有機イメージセンサーの開発

利用者氏名 : ○中田真秀\*

島崎智実\*\*

理研での所属研究室名 :

\* 情報基盤センター 計算工学応用開発ユニット

\*\* AICS 量子系分子科学研究チーム

## 1. はじめに

量子化学はここ 50 年、進展し、非常に精度の高い計算が身近なコンピュータで行えるようになった。そのため、現在化学において物質の物性、外場による応答、安定性、反応性などを議論するため、量子化学計算は必要不可欠なツールと成っている。しかしながら、「ある物性、機能を持った分子を設計したい」となると量子化学計算は途端に役に立たなくなる。この理由は、分子を知れば、(量子化学計算によって)分子の性質が解るというのを「順問題」とすると、ある性質を持った分子を創出したいという「逆問題」を解いているからである。我々はデータベース構築と、機械学習によるハイスループットスクリーニングでこの逆問題に挑む。そして、有機イメージセンサーの開発に貢献できることを目標にしている。

## 2. 具体的な利用内容、計算方法

前年度に引き続き、GAMESS[2]およびFirefly[3]を用いてPubchem[1]に登録されている分子の計算を行っている。

1. Pubchem[1]のデータのダウンロードを行う。
2. Isomeric SMILES 表記からOpenBABEL[4]を使い分子の初期座標を求めた。
3. GAMESS を用い、経験的手法である PM3 を用い、構造最適化を行った。
4. この構造を用い、ST0-6G 基底、Hartree-Fock 法で構造最適化を行った。
5. さらに、Firefly[3]を用い、6-31G(d)/B3LYP で構造最適化を行った。
6. 収束パラメータの若干の違いがあるため再び GAMESS で構造最適化を行った。
7. 最後に収束した構造からインプットファイルを作成しなおし、GAMESS で構造最適化を行った。これはインプットファイルに最適化された構造を入力し、その構造が最適化されている

かをチェックするためである。

8. 最適化された構造を用い 6-31G+(d)/TD-B3LYP で励起状態を 10 個求めた。

9. 分子量が 500 以上のもの、およびいくつかの分子の混合系については計算を除外している。

10.有機イメージセンサー[5]を作成するには、物質の HOMO-LUMO ギャップを見積もること重要である。データベースから分子の SMILES 表記から機械学習により HOMO-LUMO ギャップの予想を行った。結果は現在投降中である。

**Table 1: RMSEs (eV) between exact and predicted values of the HOMO-LUMO gap for 550,000 molecules. The predictors are supervised (trained) by 50,000 molecules, and molecular fingerprints are employed for feature vectors.**

	SVM <sup>+</sup>	SVM <sup>+</sup>	SVM <sup>+</sup>	ridge <sup>+</sup>	ridge <sup>+</sup>	ridge <sup>+</sup>
	(RBF) <sup>+</sup>	(poly2) <sup>+</sup>	(poly3) <sup>+</sup>	(RBF) <sup>+</sup>	(poly2) <sup>+</sup>	(poly3) <sup>+</sup>
MACCS <sup>+</sup>	0.490 <sup>+</sup>	0.541 <sup>+</sup>	0.674 <sup>+</sup>	0.497 <sup>+</sup>	0.511 <sup>+</sup>	0.481 <sup>+</sup>
Topological_512 <sup>+</sup>	0.490 <sup>+</sup>	0.530 <sup>+</sup>	0.593 <sup>+</sup>	0.504 <sup>+</sup>	0.518 <sup>+</sup>	0.471 <sup>+</sup>
Topological_1024 <sup>+</sup>	0.468 <sup>+</sup>	0.542 <sup>+</sup>	0.695 <sup>+</sup>	0.478 <sup>+</sup>	0.493 <sup>+</sup>	0.458 <sup>+</sup>
Topological_2048 <sup>+</sup>	0.463 <sup>+</sup>	0.607 <sup>+</sup>	0.882 <sup>+</sup>	0.467 <sup>+</sup>	0.480 <sup>+</sup>	0.462 <sup>+</sup>
Topological_1024+MACCS <sup>+</sup>	0.425 <sup>+</sup>	0.507 <sup>+</sup>	0.668 <sup>+</sup>	0.439 <sup>+</sup>	0.450 <sup>+</sup>	0.420 <sup>+</sup>

- [1] [Bolton E, Wang Y, Thiessen PA, Bryant SH, Chapter 12 IN Annual Reports in Computational Chemistry, Volume 4, American Chemical Society, Washington, DC, 2008 Apr.](#) [2] [M.S.Gordon, M.W.Schmidt pp. 1167-1189, in "Theory and Applications of Computational Chemistry: the first forty years" C.E.Dykstra, G.Frenking, K.S.Kim, G.E.Scuseria \(editors\), Elsevier, Amsterdam, 2005.](#) [3] Alex A. Granovsky, Firefly version 8.0 [4] N M O'Boyle, M Banck, C A James, C Morley, T Vandermeersch, and G R Hutchison. J. Cheminf. (2011), 3, 33. [5] NHK 技研 R&D, vol. 132, p32, 2012

平成 29 年度 利用研究成果リスト

**【論文、学会報告・雑誌などの論文発表】**

なし

**【国際会議などの予稿集、proceeding】**

なし

**【国際会議、学会などでの口頭発表】**

「PubChemQC プロジェクト:分子データベース構築と機械学習による電子構造の推定」第 11 回分子科学討論会@仙台 (東北大学) 2017/9/17

「PubChemQC プロジェクト:分子データベース構築と機械学習」2017/10/21 日本コンピュータ化学会くまもと県民交流館パレア

**【その他 (プレスリリース、学術会議以外の一般向けの講演など)】**

SudyCamp 2017 長岡研 上諏訪 2017/8/23