

力場の原子・結合タイプをラベルに用いた化学物質グラフの機械学習 Machine learning of graph for chemical substances using atom/bond types in force fields

富士通研 °實宝 秀幸, 菊地 亮太, 松尾 達, 福田 大輔, 松浦 東, 大淵 真理

Fujitsu Labs. Ltd., °Hideyuki Jippo, Ryota Kikuchi, Tatsuru Matsuo, Daisuke Fukuda, Azuma

Matsuura, and Mari Ohfuchi

E-mail: jippo.hideyuki@fujitsu.com

既知の化学物質における分子構造と特性の関係を学習し、データベースからのスクリーニングや未知の物質の特性予測を行う様々な機械学習手法が研究開発されている。化学物質の分子構造は、原子と原子間結合とがそれぞれノードとエッジに相当するグラフで表すことができる。従来は、元素記号(AS)と結合次数がそれぞれノードとエッジのラベルとして用いられてきた。本研究では、AM1-Bond Charge Correction (AM1-BCC) 電荷[1]の力場で定義される原子・結合タイプや General Amber Force Field (GAFF) [2]で定義される原子タイプをラベルに用いたグラフを提案し、生物活性有無の機械学習分類精度を検証した。データセットは、低分子化合物データベース PubChem AID778 [3] からランダムにサンプルした。機械学習分類手法には、サポートベクトルマシン(SVM)と組み合わせた 3 種類のグラフカーネル法と富士通研が開発した Deep Tensor [4] を用いた。

図 1 に 1 例として、1-(4-Ethoxy-benzenesulfonyl)-azepane の分子構造と GAFF の原子タイプをノードラベルとしたグラフを示した。図 2 は、Weisfeiler-Lehman (WL)カーネル+SVM で分類した場合の 5 分割交差検証の平均正解率である。AM1-BCC, GAFF の原子タイプをノードラベルとして用いることで、従来の AS ノードラベルよりも正解率が向上することがわかった。WL 以外のグラフカーネル手法や Deep Tensor でも同様の結果が得られている。このラベル拡張は、今回の機械学習の手法、また生物活性予測に限らず、化学物質の様々な機械学習に有効な技術であると考えられる。

[1] A. Jakalian et al., J. Comput. Chem. 23, 1623 (2002).

[2] J. Wang et al., J. Comput. Chem. 25, 1157 (2004).

[3] <https://pubchem.ncbi.nlm.nih.gov/bioassay/778>

[4] K. Maruhashi et al., Proc. 32nd AAAI Conf. Artif. Intell., 3770 (2018).

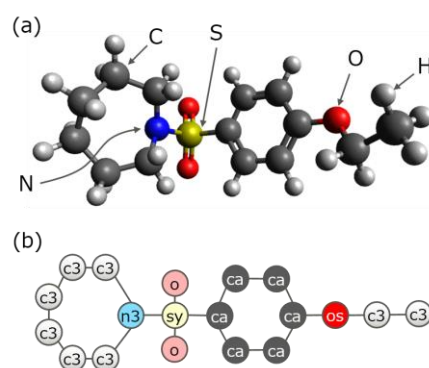


図 1: (a)分子構造と(b)GAFF の原子タイプをノードラベルとするグラフの例。

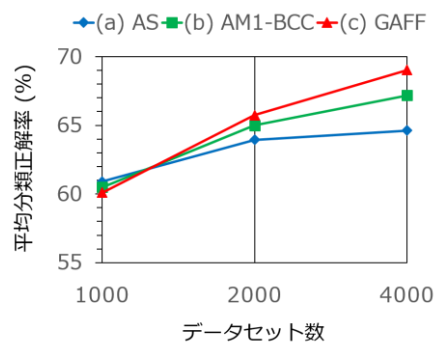


図 2: WL カーネル+SVM による平均分類正解率