

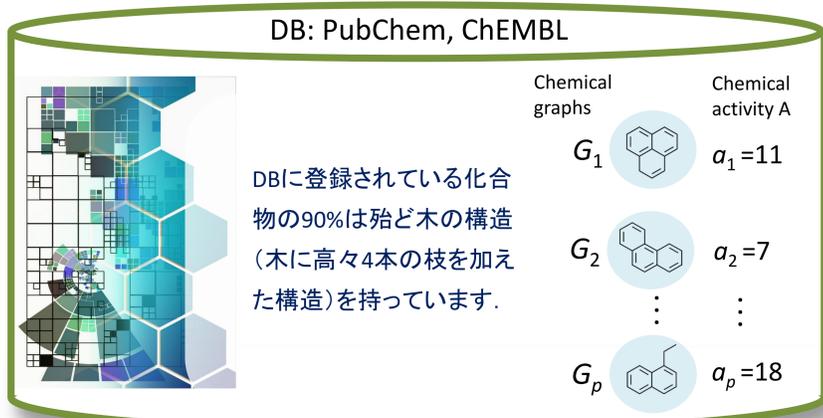
# 機械学習と離散最適化に基づく新規物質設計技術

京大化学研究所 阿久津研究室との共同プロジェクト

ケモインフォマティクスのテーマとして、有用な化学的活性をもつ新規化学物質を計算機により探し出す技術を開発しています。

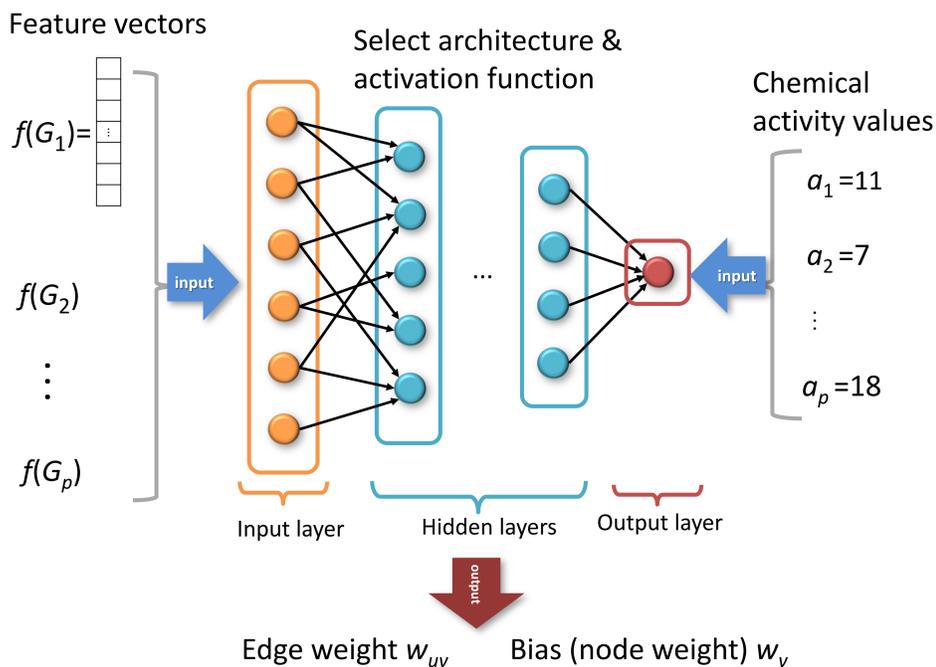
1

まず、目的とする化学的活性A(生成熱, 疎水性, 薬効等)を定め、化合物DBから化学グラフ  $G_1, G_2, \dots, G_p$  とこれらの活性値  $a_1, a_2, \dots, a_p$  を取り出します。



3

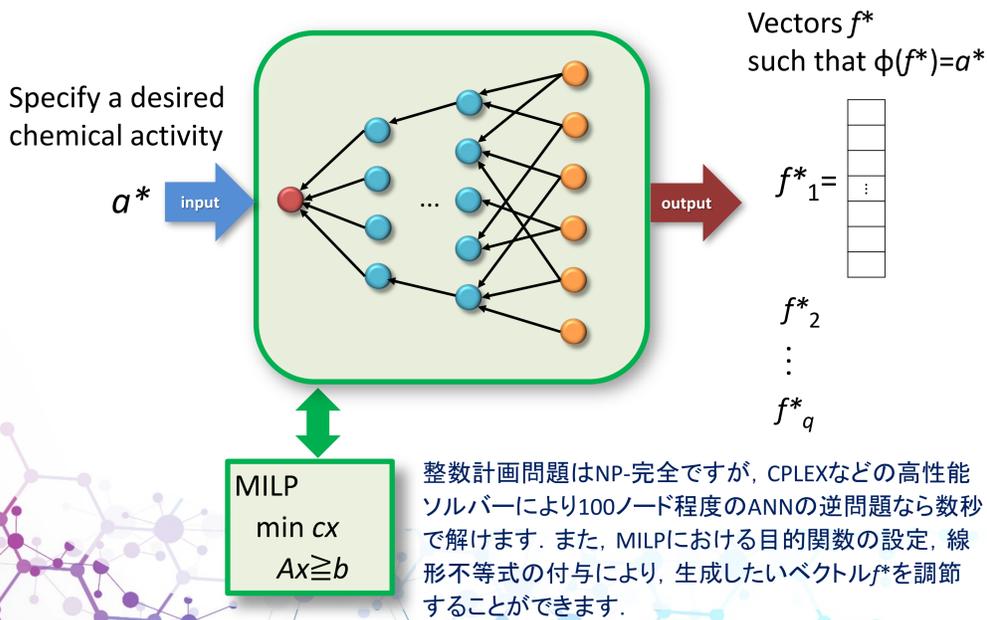
特徴ベクトルと活性値の組  $(f(G_1), a_1), (f(G_2), a_2), \dots, (f(G_p), a_p)$  を教師データとして、ニューラルネットワーク(ANN)で学習させ、特徴ベクトルと活性値の関係が最もうまく説明されるようなアーキテクチャ上の枝重みやノード上のバイアスを決めます。



5

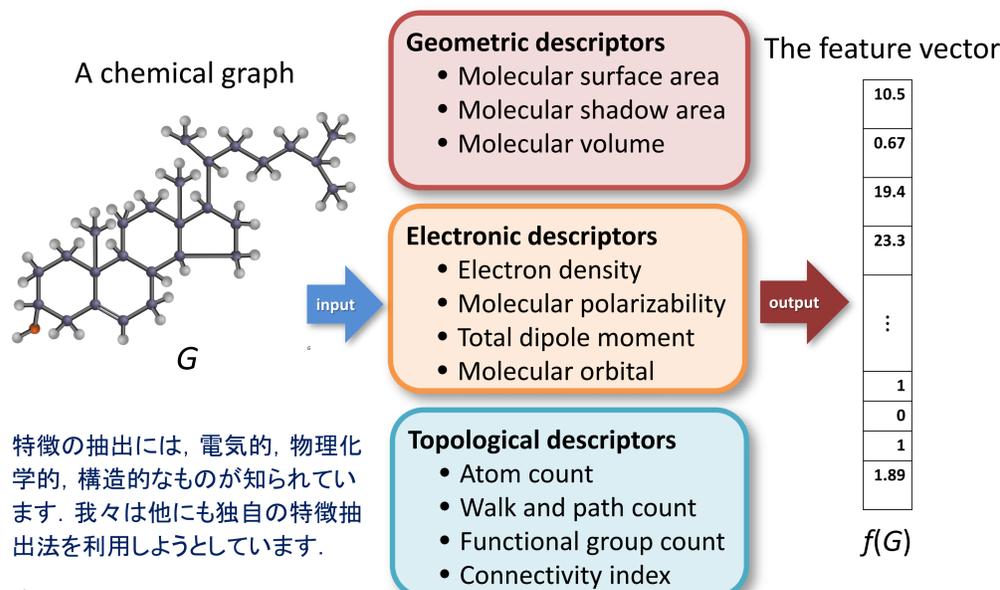
このためには、まず、指定の値  $a^*$  を予測値  $\phi(f^*)$  として出力されるような特徴ベクトル  $f^*$  を求める『逆問題』を解かなくてはなりません。我々は最近、これが混合整数計画問題 (Mixed Integer Linear Programming Problem) として解けることを発見しました。

T. Akutsu and H. Nagamochi, A Mixed Integer Linear Programming Formulation to Artificial Neural Networks, Technical Report 2019-001. <http://www.amp.i.kyoto-u.ac.jp/tecprep/index.html>



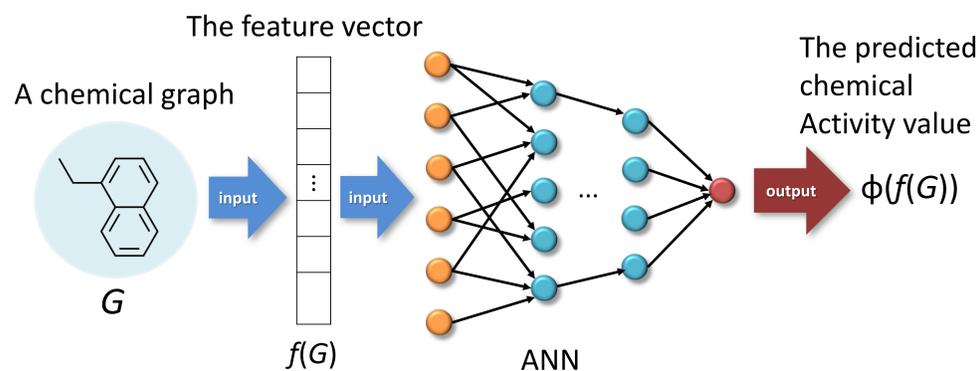
2

次に、各グラフ  $G_i$  の構造から活性Aと関係のありそうな特徴(記述子, descriptor)をいくつか抽出・算出し、これらを成分とする特徴ベクトル (feature vector)  $f(G_i)$  を用意します。



4

得られたANNを用いれば、活性の知られていない化合物  $G$  に対する活性Aの予測値  $\phi(f(G))$  が得られます。



通常は、ここまでが機械学習による従来の問題解決ですが、このプロジェクトでは、逆に、所望の活性値を入力として、この活性値を有するような化合物を予測関数  $\phi$  の原像として構築する方法を開発しています。

6

最後に、作られた特徴ベクトル  $f^*$  に対して、 $f(G^*) = f^*$  となるようなグラフ  $G^*$  を列挙します。このためのアルゴリズムは分枝限定法 (branch-and-bound) や動的計画法 (dynamic programming) に基づき設計します。

