水溶解度の推測プログラム（DNN、regression）の使い方

概要

公開されている化学物質1290個のsmilesと水溶解度（logS、mol/Lのlog10）のデータを用い、smilesの分かっている化学物質の水溶解度を推測するプログラム、solubility\_morganfp\_regression.pyのgoogle colab上での使用法について解説する。

ファイルの説明

solubility\_morganfp\_regression.py；学習データ（water\_solubility\_train.txt）のsmliles（化学式を表す記号）からモーガンフィンガープリントを計算(RDKit使用)し、それを特徴量として、AIモデルを求め、次にテストデータ（water\_solubility\_test.txt）内の化学物質の水溶解度を推測し、正解値と比較するプログラム。

1. Google Colabを立ち上げる。
2. 学習データ（water\_solubility\_train.txt）とテストデータ（water\_solubility\_test.txt）をgoogle colabのファイルの所にコピーする。
3. solubility\_morganfp\_regression.pyをメモ帳（他でもOK）で開き、内容をコピーし、google colabのセル内に貼り付ける。
4. ラン記号をONする

計算結果を水溶解度推測結果.docxに示す。また、google colab上には水溶解度の真値と推測値の表（out.csv）が表示される。良い推測ができていると思う（下図参照）。

グラフ, 散布図

自動的に生成された説明

＊RDKitは、コンピュータで化合物情報を扱う代表的なオープンソースのライブラリーで、モーガンフィンガープリントはこれを使用して計算しています。