ドッキングシミュレーションおよび
機械学習を活用した医薬品化学構造の設計

May. 26, 2016
ディスカッション

酒井研究室　　宮崎 大輝

学習用データおよびテストデータのランダム抽出

logS_data_set_2D_original.sdf　から
構造記述子データを計算した構造に対するlogSの値を抽出
→logS_data_set.csv

logS_mcd.csvからa個の学習用データの構造記述子データをランダムに取り出し
→ x.csv
logS_data_set.csvから説明変数と対応するa個のlogSデータを取り出し
→ y.csv
1170-a個のテストデータの構造記述子データを取り出し
→xeval.csv

学習用データの取り方、及び学習用データの個数を変え、複数回最適候補の探索

GP法を用いた予測モデル

予測性能の確認

評価値：期待値、yの制約条件：最大化　探索候補数：１　として最適候補の探索

学習用データ 20個
(テストデータ1150個)

試行回	y予測値	y観測値	誤差
1	-1.21928608	-0.85	-0.37
2	-0.46826815	-7.32	6.9
3	-0.23908013	-1.85	1.6
4	-1.73676148	-5.3	3.6
5	-2.10040219	-0.13	-2.0

学習用データ 50個
(テストデータ1120個)

試行回	y予測値	y観測値	誤差
1	1.3722739	1.34	0.032
2	1.51543931	0.58	0.94
3	-0.45309902	-0.46	0.0069
4	0.15821182	0.28	-0.12
5	0.70928619	-1.08	1.8

- **全体的に高いlogS予測値が得られた**
- **学習用データの個数を増やした方が予測誤差は減る？**

今後の予定

- さらに抽出するデータ個数および抽出の範囲を変えての探索
- 探索候補数など他のパラメタを変えての探索
- ドッキングシミュレーション結果を用いた探索
- GP法の理解