

インターンシップ報告書

工学系研究科 物理工学専攻 D1 安田寛徳 (求研究室)

研修先：日本ゼオン株式会社

本インターンシップでは高 CNT 成長性を示す炭化水素の機械学習による探索という題目でインターンシップを行った。

【研修内容】

・期間：2020.10.5.-2020.12.10

・業務は主にオンラインで実施され、数回勤務地である総合開発センター(川崎市)へ出勤し、議論や研究所の見学を行った。また、最終日には成果報告会を行った。

【課題の概要】

工業的にも注目されるカーボンナノチューブは触媒上に炭化水素を吸着させて合成されるが、その成長性は吸着する炭化水素によって大きく変化する。本研究課題は、より多くの収量をえるのに最適な炭化水素の構造を数値計算によって探索することを目的とする。

具体的には、触媒上で様々な炭化水素の成長についてシミュレーションを行い、それらのデータから最適な炭化水素を探索する。

まず、第一原理計算によって炭化水素の反応のデータの追加を行う。そして、現在までに得られているデータと合わせた訓練データをもとに、回帰分析等を用いた炭化水素の推定を行い、推定の精度の向上を目指す。

しかし、炭化水素の構造のバリエーションは炭素数に応じて指数関数的に増加するため、炭素数の多い炭化水素において全ての構造パターンに対する計算を行うことは計算コストの面で現実的に困難である。そこで、ベイズ最適化を用いて候補となる可能性が高い炭化水素の構造群を重点的に探索し、訓練データの計算コストの削減を試みる。ベイズ推定には自作あるいは既存のパッケージを使用することを検討している。

【手法】

まず、オープンソースパッケージ Phase0 を用い、現在までに第一原理計算で得られたデータの確認と検証を行った。また、ベイズ推定においては、モンテカルロ法を用いたベイズ推定パッケージ Stan を用いてカーネル法による解析を実装した(図1)。具体的には複数種のカーネル関数でモンテカルロサンプル数を変えて交差検証を行い、R2 値を比較することでデータに適した確率モデルを最適化した。

この際、入力として炭化水素の2次元特徴量を主成分分析で次元削減したものを使用した。また、統計的な外れ値とみなされるデータは除外した。そして、それを用いたベイズ最適化を行い、吸着性の良い炭化水素を推定した。さらに、それらの炭化水素に

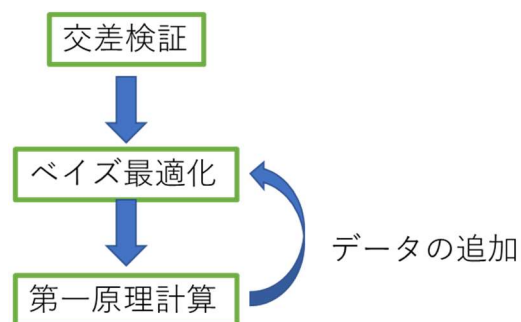


図 1：課題解決の流れ

対して Phase0 で第一原理計算を行い、学習データにその結果を追加した。本インターンシップではこのプロセスを数回反復した。

【結果】

ここではベンチマークとして行った、ボストン近郊での住宅価格のデータに対する推定結果を紹介する。カーネル関数としては 3 次のマテルン関数 $K(x, x') =$

$$p_1^2 \left(1 + \sqrt{3} \frac{d(x, x')}{p_2} \right) e^{-\frac{d(x, x')}{p_2}}$$

を用い、モンテカルロサンプル数は 2000 とした。R2 値は 0.89 と非常によい。図 2 はこのときの予測値を実際のデータと比較したものであるが、確かに両者はよく一致している。また、図 3 はこれらの予測値と実際のデータに 95%信頼区間を付けてプロットしたものであるが、実際のデータが概ね 95%信頼区間内に収まっていた。研修ではこのコードを用いて、炭化水素のデータに対して同様の計算を行った。

【感想】

本インターンシップでは機械学習を題材に企業において計算機科学が、実験研究や製造プロセスの改善のための予測システムとして熱心に取り組みられている現場の一端を体験できたと思う。それに加えて私自身経験のなかった機械学習の知識を勉強し、実際に活用することで機械学習についての理解も深めることができた。また、様々な研究チームの方と交流する機会を設けていただき、自分のやっている解析と実際の工程の関係性を実感しながら業務に励めた。その際、頻繁に協力しているチームと議論し、すり合わせを行っている様子が印象的であった。有機化学系の分野は無機化学の分野に近い自分とは専門性の違いがあり、その中でコンセンサスを取りながら課題を進めるコミュニケーション能力の重要性を再認識した。大変実りのあるインターンシップであったと思う。

【謝辞】 まず、コロナウイルスで受け入れが困難な中、基盤技術研究所の皆様が私を受

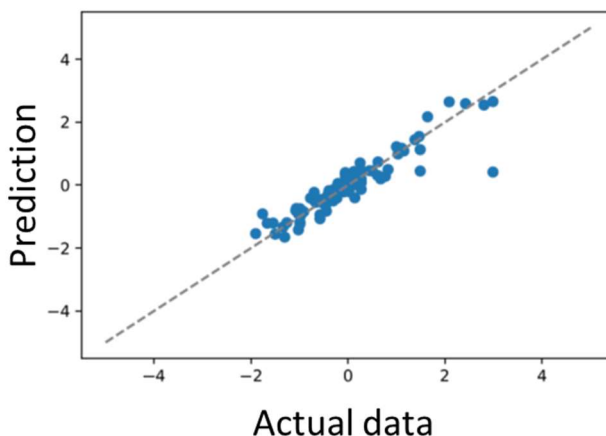


図 2: ボストン近郊の住宅価格データに対するベイズ推定の予測値と実測値の比較。点線は縦軸=横軸の直線。

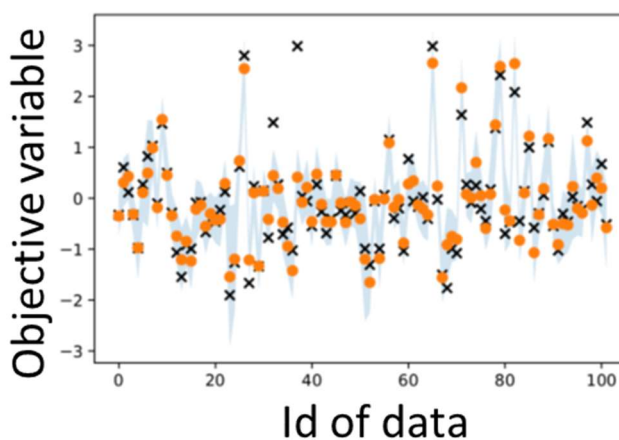


図 3: 各データの予測値(x)、実測値(o)と 95% 信頼区間のプロット

け入れていただいたことに感謝いたします。計算チームの皆さまにはお忙しい中、テーマ設定や議論、データ処理など、多くのお時間を割いていただきました。また、座談会や研究紹介など、研究所の方々との交流の場を設けていただきました。実験チームの方には研究所の案内などをしていただき、また実験の立場からご意見をいただきました。