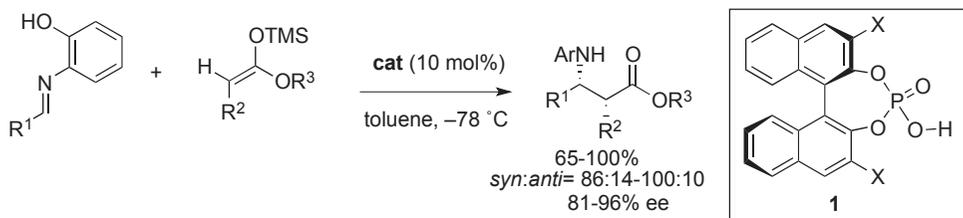


## 機械学習を利用した分子触媒開発のための予備的研究

理学部教授 秋山 隆彦  
 計算機センター教授 久保山 哲二  
 理学部助教 宮川 雅道

機械学習 (machine learning) とは、人工知能の要素技術の一つであり、人間が行っている学習能力と同等の機能をコンピューターで実現しようとする技術・手法の一つである。近年、化学の分野でも機械学習が注目を集めており、機能性材料の開発等に応用され、開発の効率化に貢献していることが明らかになっている。例えば、アメリカ化学会発行の機関紙である *Chemical & Engineering News* でも 2017 年の 1 月第 4 週発行の号において、"New directions for machine learning: Algorithms that have changed daily life move into the lab" のタイトルで、機械学習の化学への応用についての特集が生まれ、化学産業において、機械学習を用いた材料開発の例などが紹介されている<sup>1)</sup>。また、ごく最近、"Is machine learning overhyped?" のタイトルで、機械学習の現状に関する記事も掲載されている<sup>2)</sup>。

一方、化学の分野においては、優れた不斉触媒を開発し、効率の高い触媒反応を開発することは、医薬品等の生理活性物質の効率的な合成につながることから、重要な課題の一つであり、幅広い種類の不斉触媒が開発され、実験室レベルから大量合成までに用いられている。一般に優れた不斉触媒反応を開発するためには、試行錯誤により、膨大な数の実験を行うことが必要不可欠であるのが現状である。



Scheme 1. エナンチオ選択的 Mannich 型反応

秋山研究室では、(R)-BINOL 由来のキラルリン酸 **1** を合成し、キラルリン酸がキラルプレンステッド酸として優れた不斉触媒能を示し、Mannich 型反応を効率良く触媒することを 2004 年に見出した<sup>3)</sup> (Scheme 1)。その後、キラルリン酸 **1** は幅広い種類の不斉触媒反応に有効であることを報告してきた。また、世界中で膨大な数のキラルリン酸 **1** を

用いた不斉触媒反応が報告されている<sup>4)</sup>。しかし、キラルリン酸を用いた不斉触媒反応を開発するためには、様々な条件を検討する必要がある。例えば、先の Mannich 型反応を例にとると、例えば、以下の3つの条件を最適化しなければならない。

変更する点：(1) リン酸の3,3'-位の置換基，(2) 反応溶媒，(3) 反応温度

その条件に対し、以下の収率および光学純度ともにより良い結果を得る必要がある。

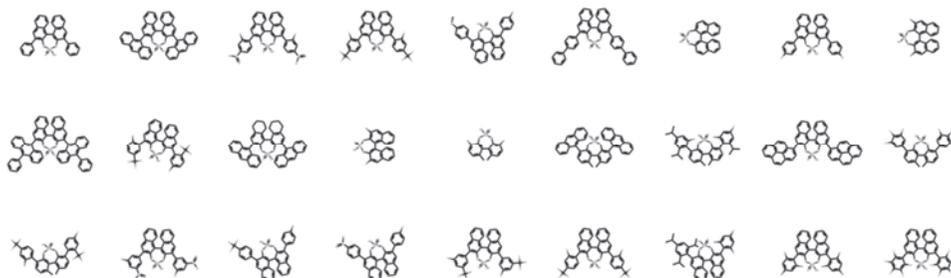
結果：(a) 生成物の収率 (%)，(b) 生成物の光学純度 (% ee)

実際の研究においては、一つ一つ実験を行い、これらの条件を少しずつ変更し、最適条件を見出すことが必要であり、数多くの実験、長い時間を要する。近年、コンピューターの性能が向上したために、理論化学計算により、遷移状態を計算し、選択性の発現する理由を解明することが容易になってきた。既存の反応の遷移状態を明らかにし、置換基効果についての知見を得ることできるものの、新規な反応における最適触媒を予測することは困難であるのが現状である。新たな反応を開発する際に、何らかの手法により、最適触媒ならびに諸条件の最適化が可能になれば、反応開発が大きく効率化されることが期待される。

本研究では、これまでに蓄積された実験条件と結果のデータをデータ分析することにより、優れた分子触媒を開発するための条件探索を効率的に探索できるのではないかと考え、そのためのデータ形式の調査やデータの前処理に関する予備的な調査を行った。

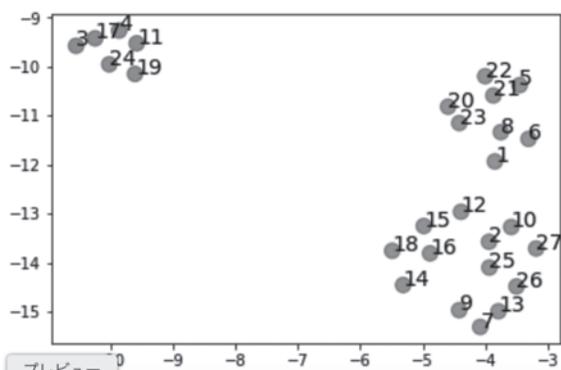
まず、数多くのキラルリン酸誘導体の化学構造を、化学で汎用されている構造式描画ソフトである Perkin Elmer 社製 ChemDraw の形式で蓄積している。ChemDraw により作成した化学構造は、ChemDraw Binary Format (拡張子 .cdx)により保存されており、これらのデータを、コンピューター上での処理に適用するためには、描画データを適切な形式に変換する必要がある。ChemDraw で作成した構造式は XML により記述された CDXML 形式に変換できることから、立体構造描画の任意性により一意性はやや犠牲になるものの木構造データへの変換が可能であることを確認した。今回は、化学構造フォーマット変換ツールである Open Babel を用いて ChemDraw Binary Format のデータを変換した。

次に、機械学習の既存手法に化学構造式を適用するためには、何らかの形で計量化する必要がある。化学構造間の距離や類似度を何らかの方法で計測できれば、カーネル法を用いた分類・回帰や、クラスタリング等の様々な機械学習の手法に適用できる。今回は、Morgan/Circular FingerPrint (Radius=4, 4096bits)により化学構造をベクトル化した。以下にベクトル化した27の触媒を示す。



各々の触媒が FingerPrint により 4096 ビットのベクトルに対応しており、このベクトル間の距離を Tanimoto 距離を用いて計測した(Tanimoto 距離は、ビット列を集合表現とみなすと Jaccard 距離に等しい)。さらに、次元削減手法である UMAP (Uniform Manifold Approximation and Projection)を用いて 2 次元に各々の触媒をマッピングした。

過去の秋山研究室で行った、130 の実験データにおける反応、触媒、溶媒、温度の 4 つの特徴 (次元) を用いて、CatBoost 回帰により触媒の収率の予測を試みたが、予測性能は低かった。これは、データ量が少なすぎることで、特徴数が 4 つと少ないことに起因する。



本報告では、秋山研究室で蓄積した実験データを今後有効にデータ分析し、効率的な実験計画を立てるための形式やツールの準備、予備的な分析を行った。実際にデータ分析により有効な結果を得るためには電子化された実験データが不足しており、今後、実験条件を分析しやすい形で記録する等の手順を整え、研究室で蓄積した大量の実験記録を分析可能なデータに変換していく予定である。

### 参考文献

- (1) Kemsley, J.; Wilson, E. K.; Mullin, R. *Chem. & Eng. News* **2017**, *95*(4), 28.
- (2) Lemonick, S. *Chem. & Eng. News* **2018**, *96*(34), 16.
- (3) Akiyama, T.; Itoh, J.; Yokota, K.; Fuchibe, K. *Angew. Chem. Int. Ed.* **2004**, *43*, 1566-1568.  
Akiyama, T. *Chem. Rev.* **2007**, *107*, 5744-5758.
- (4) Parmar, D.; Sugiono, E.; Raja, S.; Rueping, M. *Chem. Rev.* **2014**, *114*, 9047-9153.
- (5) Akiyama, T.; Morita, H.; Fuchibe, K. *J. Am Chem. Soc.* **2006**, *128*, 13070-13071.