

| | |
|--------------|---|
| Title | 特微量設計に基づく機械学習を用いたメタン酸化カップリング触媒の開発 |
| Author(s) | 藤原, 綾 |
| Citation | |
| Issue Date | 2025-03 |
| Type | Thesis or Dissertation |
| Text version | none |
| URL | http://hdl.handle.net/10119/19939 |
| Rights | |
| Description | Supervisor: 谷池 俊明, 先端科学技術研究科, 博士 |

| | | | |
|---------------|---|---------------|-----|
| 氏 名 | 藤原 綾 | | |
| 学 位 の 種 類 | 博士（マテリアルサイエンス） | | |
| 学 位 記 番 号 | 博材第 603 号 | | |
| 学 位 授 与 年 月 日 | 令和 7 年 3 月 21 日 | | |
| 論 文 題 目 | Development of methane oxidative coupling catalysts with feature engineering-based machine learning | | |
| 論 文 審 査 委 員 | 谷池 俊明 | 北陸先端科学技術大学院大学 | 教授 |
| | 長尾 祐樹 | 同 | 教授 |
| | 西村 俊 | 同 | 准教授 |
| | 本郷 研太 | 同 | 准教授 |
| | 中山 哲 | 東京大学 | 教授 |

論文の内容の要旨

Materials Informatics (MI) aims to accelerate the discovery and understanding of materials through data science, especially in cases where complexity exceeds human intuition. However, there are two main challenges to implementing MI in practical applications: one is the lack of data that meets the quality, scale, and consistency required for effective machine learning, and the other is the need for advanced domain knowledge to design descriptors (numerical representations of catalysts). To address the data challenge, Taniike and his research group have developed their own datasets using high-throughput experimentation (HTE). Additionally, they introduced an automatic feature engineering (AFE) method that enables the design of descriptors without domain-specific knowledge, thereby addressing the descriptor challenge. This technique generates numerous features (descriptor candidates) from the physicochemical properties of catalyst components and extracts those that are relevant for describing catalyst performance.

The oxidative coupling of methane (OCM) is a catalytic reaction that directly synthesizes ethylene (C₂H₄) from methane (CH₄). This process is significantly more energy-efficient than conventional synthesis routes via syngas. However, due to the higher stability of CH₄ compared to C₂H₄, achieving high yields remains challenging, necessitating the development of high-performance catalysts. Taniike and his group have accumulated a dataset containing performance data for over 600 catalysts in OCM using their HTE system. Using this data as the training dataset, they have expanded the scope of catalyst exploration with machine learning techniques, such as decision tree analysis and support vector regression (SVR). These approaches, however, adopted one-hot encoding to represent catalyst compositions, where a catalyst component is represented as either 0 (element x is absent) or 1 (element x is present). The limitation of this method is that it treats catalyst compositions as symbols without providing insight. While researchers may recognize that catalysts A and B share common physical properties, the machine learning model perceives only common symbols. For efficient catalyst exploration, it would be ideal to define some physical features as a descriptor, however, achieving this in such a complex catalytic reaction is highly challenging. This illustrates a core challenge in MI: the need for a machine learning framework capable of generating hypotheses without requiring prior knowledge of the target system, particularly for the vast exploration of candidate materials.

From the perspective that “researchers derive insights from multiple observations,” when observational data is limited, it becomes difficult to reject competing hypotheses. Thus, researchers conduct repeated control experiments

to refine their hypotheses and arrive at accurate conclusions. This hypothesis refinement process is heavily reliant on a researcher's intuition, impacting both the speed and quality of research outcomes. If this process could be integrated into a machine learning framework and streamlined with high-throughput experimentation, it could establish a versatile research framework applicable to a wide range of experimental research.

In this thesis, a refinement process is simulated using an adaptive catalyst design cycle that combines AFE, farthest point sampling (FPS), and HTE, targeting OCM catalysts. The study aims to develop and validate a robust machine learning model, demonstrating its utility in understanding catalyst design principles and supporting efficient catalyst development.

In **Chapter 2**, an efficient catalyst discovery process combining data-driven AFE, active learning, and HTE was demonstrated targeting BaO-supported catalysts for OCM. Through the refinement of feature space across diverse catalysts via AFE and HTE, predictive accuracy was enhanced, leading to the discovery of high-performance catalysts with C₂ yields exceeding 15%.

Building on this methodology, **Chapter 3** expands the scope to multiple catalyst supports, including BaO, CaO, La₂O₃, TiO₂, and ZrO₂. The analysis revealed two main patterns: one where a single element, such as La on BaO, dominated performance, achieving high yields without additional elements; and another, exemplified by CaO, where a combination of multiple elements, particularly alkaline earth metals with Cs, was essential. These findings highlighted distinctive design rules for each support.

In **Chapter 4**, active learning efficiency was explored to achieve high-accuracy learning with smaller datasets. Additionally, I explored the possibility of applying design hypotheses derived from known supports to new supports. We tested whether design hypotheses from the five supports in Chapter 3 could guide predictions for Y₂O₃ supports. This approach confirmed the transferability of design insights across supports.

The catalysts newly developed in Chapters 2, 3, and 4 are synthesized in **Chapter 5**, which highlights the high-performance catalysts created throughout the thesis.

In summary, this thesis presents an original adaptive catalyst design cycle with recursive feature engineering, proving the utility of the resulting machine learning models in understanding catalyst design heuristics and enhancing catalyst development. By exploring a broad range of compositions for OCM catalysts, this approach successfully identified high-performance catalysts and valuable design heuristics.

Keywords: Catalyst informatics, Machine learning, High-throughput experimentation, Oxidative coupling of methane, Descriptor design technology

論文審査の結果の要旨

本論文は、既知の因果関係や人間の洞察を必要とせずに成立するデータ駆動型の触媒開発方法の確立に関する研究成果をまとめたものである。

触媒は、化学プロセスの 8 割を担い、様々な製品を介して現代の物質社会を支える根幹的な材料である。しかし、触媒の構造や反応機構は極めて複雑であるため、因果関係の全貌の把握が難しく、その研究開発は実験的な試行錯誤に依存してきた。そこで、因果関係ではなく相関関係に基づくデータ駆動型の触媒開発が注目されてきた。しかし、これには、量と質を兼ね備えた触媒データの不在に加え、何より、複雑な触媒を適確に表現するための記述子を研究者が機械に与える必要がある、という課題があった。そこで本研究では、ハイスループット実験（HTE）と自動記述子設計（AFE、データから機械に記述子を自動生成させる技術）を基盤として、能動学習を通して、既知の因果関係や人間の洞察を一切必要とせずに、汎化能力に優れた高精度な学習モデルを獲得する方法論を確立した。メタンをエチレンへと直接的に転換する、メタン酸化的カップリング反応（OCM）を対象に用い、方法論の有用性を実証した。

第 2 章では、BaO 担持 OCM 触媒を対象に能動学習を実施し、自動生成された 8 個の記述子から成る高精度な学習モデルを獲得した。このモデルを詳細に解析することで、BaO 担持触媒の設計則を明らかにした。得られた学習モデルを活用し、従来よりも遥かに高い効率で多様な高性能触媒を狙い撃つことにも成功した。

第 3 章では、研究対象を BaO、CaO、La₂O₃、TiO₂、ZrO₂ の 5 種類の担体に拡大した。これまでの触媒研究では、これらは別個の研究として、それぞれ特有の設計則に基づいて開発されてきた。そのため比較対象が限られており、ある系で得られた知識を他の系にそのまま応用することは困難であった。しかし、これらの系における学習モデルを別々に構築し、同じ元素組み合わせを担持した際の性能を比較することで、異なる系との類似度を評価することが可能となった。また、予測された触媒性能と担持元素の相関を調べた結果、担体特有の設計則に加え、後周期遷移金属の阻害効果やアルカリ（土類）金属元素の促進効果など、異なる系の間に共通する触媒設計則が存在することが明らかになった。

第 4 章では、Y₂O₃ 担持触媒を対象に、より少ない能動学習サイクルで汎化能力に優れた学習モデルを構築する方法を考案した。他の系で導いたモデル（知識）を転用すること、及び、未知の触媒に対して物理的でない予測を示し得るモデルを排除するフィルターを併用することで、これまでの半分以下の触媒数で同等の精度を持つモデルを得ることに成功した。

第 5 章では、本論文を通して発見された触媒を概観した。新たに発見された高性能触媒は新規性が高いものが多く、方法論の有効性が示された。

以上、本論文では、多様な触媒系に適用可能なデータ駆動型の触媒開発法を提案し、その有効性を大規模な実験で実証した。触媒分野への波及効果は大きく、実用性も高い。以上の成果から、博士（マテリアルサイエンス）の学位論文として十分価値あるものと認めた。