

Title	データサイエンスを活かした固体触媒開発: 固体触媒インフォマティクスの実践
Author(s)	西村, 俊
Citation	光機能材料研究会報 光触媒(データサイエンス研究の基礎と最新技術・応用展開), 75: 29-38
Issue Date	2024-09-13
Type	Journal Article
Text version	author
URL	http://hdl.handle.net/10119/19340
Rights	本著作物は技術教育出版の許可のもとに掲載するものです。This material is posted here with permission of Gijyutukyoku. Copyright (C) 2024 技術教育出版. 西村俊, 光機能材料研究会発行「会報光触媒75号」, 2024, pp 29-38.
Description	

データサイエンスを活かした固体触媒開発：固体触媒インフォマティクスの実践

西村 俊¹

¹北陸先端科学技術大学院大学 先端科学技術研究科 石川県能美市旭台 1-1

機械学習手法やデータマネジメント技術を活用した『データ駆動型の材料開発 (X-インフォマティクス)』が様々な研究分野に展開されている。触媒化学の分野でも従来型のアプローチと比較してより効率的・簡便に所望の性能を発揮できる触媒材料開発ないしはプロセス最適化を実現可能な「触媒インフォマティクス」の実装が期待されている¹⁻³⁾。究極的にはロボットを展開した実験自動化によるデータ蓄積と、学習機(ないしは AI) がデータに基づいた新たな材料・プロセス設計を展開することで、経験豊富な研究者の知識や技能を介さない「よりスマート」な材料開発のオートメーション化が展望されている。その一方で、データ駆動型のアプローチはデータの大多数の傾向を捉えているだけで真新しい予測を得られないといった報告⁴⁾など、手法への期待だけではなく実際に材料探索へ用いた際の技術的限界点やその打開策にも目を向ける必要がある。本稿では、2017 年から JST-CREST 事業の支援の下で推進した『固体触媒インフォマティクスプロジェクト (代表:高橋啓介)』活動の一端について紹介する⁵⁾。

1982 年に報告されたメタン酸化カップリング反応 (OCM: Oxidative Coupling of Methane) は、天然ガスの主成分であるメタンを直接カップリングさせることで主に C2 化合物 (エチレン、エタン) を与えるプロセスであり^{6,7)}、ナフサのクラッキングを経由しない C2 生産プロセスとしてその商業化が期待されてきた。40 年余りの研究の中で、Li/MgO や NaMnW/SiO₂ などの有望な触媒の発見、スタートアップによるプラント稼働の試みが行われているが、未だに採算性を担保した商業プロセスの確立が難しい反応系である。実際に C2 収率 30%以上ないし C2 選択率 80%以上 (収率 18-24%) が目標性能とされる中で、2010 年から 2017 年の触媒技術開発力の滞留も指摘されている状況にあった⁸⁾。そこでこれまでとは異なるデータ駆動型のアプローチを展開することで、新しい OCM 触媒開発への活路が見いだせないかと考えた。また OCM 触媒開発は達成目標値が明確であり、更に過去の OCM 触媒開発に関する文献データを集約した総説⁹⁾も出されていたため、自前の計算データや実験データを大規模展開するまで

の準備段階でも、ある程度まとまった量のデータにアクセス可能な状況であった。さらに基本的には 800°C 付近の高温を必要とする反応であるため、データ駆動型の触媒開発を展開する上で触媒材料記述を比較的構成元素の種類に置き換えて表現しやすい反応系でもあるといった利点も有していた。これらの点を加味して、データサイエンスを活かした固体触媒開発のケーススタディとして OCM 反応系を選定した。

まず、文献データに基づく分類器による予測¹⁰⁾と提案触媒の実験実証¹¹⁾を行った結果を紹介する。文献データ（学術論文、特許など）は、実験研究者の知識や経験に基づく従来の触媒探索アプローチにより構成されたデータセットであり、研究のムーブメント、材料トレンド・技術限界等を教示可能である。様々な研究者が試行錯誤の末に獲得した経験知の集合体であり、研究グループごとの着想に沿って多様で広範囲な元素探索の履歴を包括している。一方で、触媒調製スキームや反応評価条件・触媒機能の評価手法の多様さがデータのバイアスとなる課題もある。通常、実験研究者はその一部を参照し、自らの知識・経験に基づくひらめきや勘を頼りに新たな材料設計を行い、新しい発見や機能改善を実現している。「機械学習を用いることで網羅的に全データの傾向を読み解くことができれば、より真新しい触媒設計に繋がるのではないか？」といった視点を基に、1982 年から 2011 年の過去 30 年間に報告された 1,868 点の OCM 触

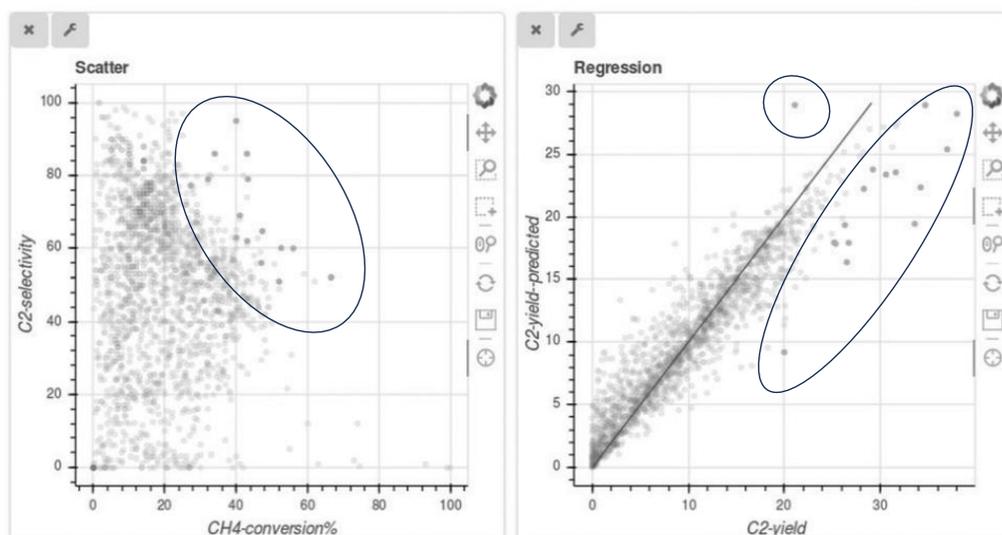


図 1 文献データ⁹⁾の C2 選択性 vs CH₄ 転化率分布（左）と RFR 回帰を実施した際の交差検証結果における外れ値の分布評価¹²⁾ ※○で示した領域が主に外れ値を示す。 [英国王立化学会（The Royal Society of Chemistry）により掲載許諾（ref.5）]

媒データ⁹⁾を学習した分類器 (Random Forest Classification) を作成し、触媒組成と反応条件のトレンドを学習させた。次いで 11 種類の記述子を基に 331,164,288 通りの仮想データ空間を設定し、C2 収率 30%超を実現可能な 2 元素の金属を担持した OCM 触媒 (M1-M2/Support) を予測させた。この方法により未探索な 56 触媒組成およびその時に望ましい反応条件を提示することができた。但し、取り扱いが難しい放射性元素や CH₄-O₂ の爆発限界に踏み込む危険な反応条件も含まれていた。提案触媒材料の有用性の検証は、NaMnW/SiO₂触媒が C2 収率 19.0%を発現できる反応条件 (一定) を用いて簡略的に実施したが、C2 収率 30%超を発現しうる可能性を有する触媒 (NaMnW/SiO₂触媒よりも有望そうな OCM 触媒) の発見には至らなかった。C2 収率の予測精度が高い学習機 (分類器) の構築が、必ずしも OCM 触媒で C2 収率が高い触媒設計に繋がるわけではないことが示される結果となった。元の文献データの傾向を再解析するために RFR (Random Forest Regression) による回帰を行った結果、外れ値のデータ群が C2 収率 30%を超える触媒データ分布に一致することが分かった (図 1)。元の文献情報を辿ると、外れ値の触媒系は特異な技術 (特殊な結晶系の合成や反応装置機構) に由来する触媒性能であることが確認できた。従って、全体のデータの傾向を掴むことを原動力とするデータ駆動型の探索では、このようなトレンドから逸脱した空間 (ここでは C2 収率 30%超) での触媒予測が実現できなかつたと結論付けた。

次に、ハイスループットスクリーニング (HTS) 実験データと文献データを組み合わせたデータセットに基づく触媒探索の結果¹³⁾について紹介する。HTS データは、単一の反応装置を用いた大規模スクリーニングにより収集される実験データで、触媒調製条件 (調製法、金属担持量など) と反応評価機構 (スキーム、評価装置設計など) が比

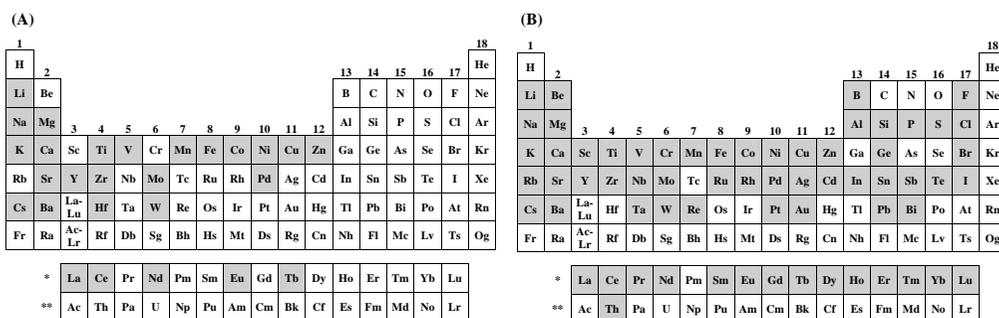


図 2 (A) HTS 実験データ¹⁵⁾および(B) 文献データ¹⁶⁾の探索範囲の周期表上での分布

Cat. Comp.	[Cat. comp. described by a one-hot encoding manner]										[Reaction condition]			[Target]	
	Na	Ce	Mn	W	Li	none	TiO ₂	SiO ₂	MgO	· · ·	Temp.	P _{CH₄}	P _{O₂}	P _{balance}	C ₂ yield
NaCeW/TiO ₂	1	1	0	1	0	0	1	0	0	· · ·	973 K	0.55	0.30	0.15	x %
LiMnW/SiO ₂	0	0	1	1	1	0	0	1	0	· · ·	1073 K	·	·	·	y %
LiNaNone/MgO	1	0	0	0	1	1	0	0	1	· · ·	1173 K	·	·	·	z %

図 3 One-hot encoding 法を用いた触媒材料記述方法および反応条件セットの例示

較的均質な条件下でデータ収集を行うことが可能である。任意のパラメーターを固定できることから比較的バイアスが少ないデータを収集可能であり、パーソナルスキルに関しても同じ研究チームでの教育プロセスを経た実験者による操作であれば不均質さの最小化が期待できる。一方で、文献データと比較すると探索範囲にはマンパワーや範囲設定時の観点に依存する制限が生じる（図 2 A に 300 種類の触媒を HTS 実験に実装した探索空間を例示する）。ここでは、共同研究チームからオープンソースとしてオンラインプラットフォーム（CADS: Catalyst Acquisition by Data Science）¹²⁾上に公開されている 300 + 59 種類の OCM 触媒に関する HTS 実験データ（40,339 点）^{14,15)}を用いた。また文献データとしては、2021 年までのデータを含む 4,759 データ¹⁶⁾を使用した（図 2 B に当該文献データの探索空間を示す）。データ前処理として、HTS 実験データ（40,339 点）では、各触媒組成に対して最大 C₂ 収率を示す 1 データのみを抽出し、文献データと統合した（HTS 触媒組成へのデータ偏重を抑制するため）。さらに、炭素収支の不確かさや 3 元素以上の金属元素を含むデータを除去し、2,842 点のデータを選別・抽出した。触媒組成の認識には One-hot encoding 法を採用し、元素の index 表に 0/1 を書き込むことで元素の有無を表現し、対応する反応条件と C₂ 収率を紐づけた学習機（SVR: Support Vector Regression）を作成した（図 3）。従来の実験研究者は自らの知識・経験に基づく仮説検証を行い、実際の結果を基に再考を行うことで目標とする新発見や性能改善を実現している。「機械学習機に自らの予測データの実証結果を追加して再学習させた時に、より高次の触媒設計に繋がるのか？」といった問いに基づき、実験検証を行った。

まず SVR(1)で提案された最大 C₂ 収率予想値 22.61%の空間の中で、18.00%以上の 36 触媒について実験検証を行った。なお、OCM 反応は反応雰囲気¹⁷⁾だけでなく評価装置設計¹⁸⁾により大きく C₂ 収率が変動することが知られている。従って、多様な反応器により収集された文献データ（ないしは実験検証と異なる反応器で収集された HTS データ）を基に C₂ 収率の発現に必要な反応雰囲気が提案されているため、実験実証段階で

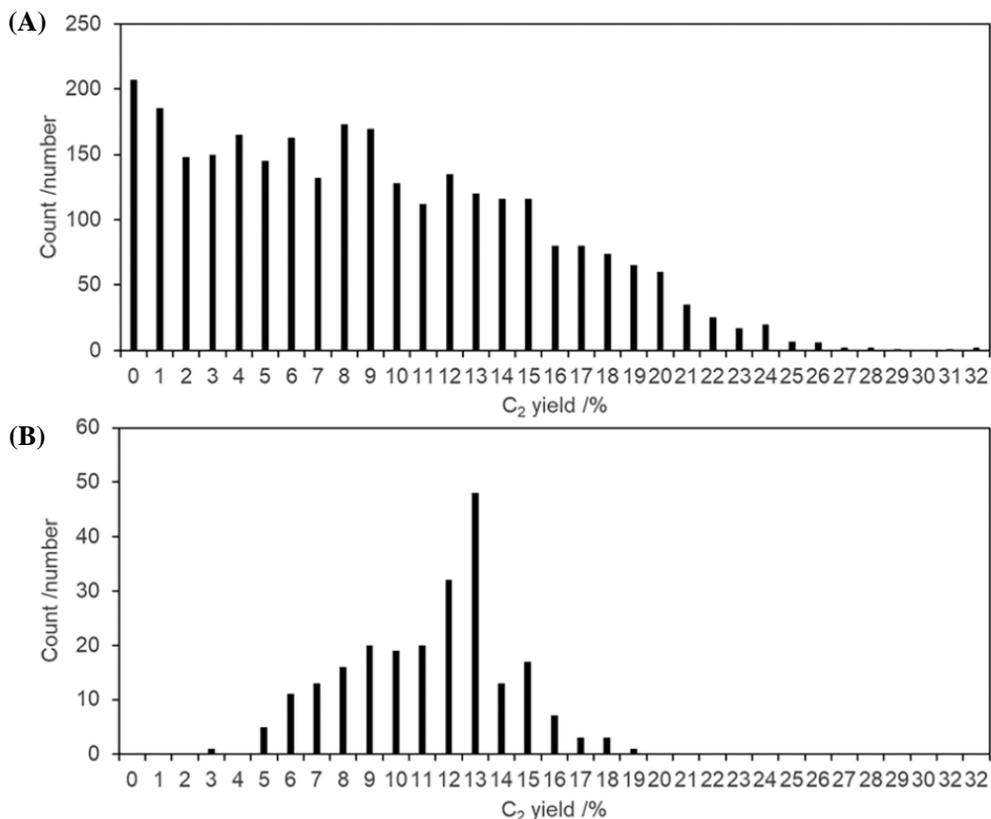


図4 (A) 元の学習データおよび (B) 実験実証を基に追加した実験データの C₂ 収率分布¹³⁾ [英国王立化学会 (The Royal Society of Chemistry) により掲載許諾 (ref.13)]

異なる装置機構の装置を用いてその反応雰囲気を実現しても予測 C₂ 収率値の再現が困難であることには注意が必要である。ここでは、提案された反応雰囲気の情報に基づき、実験実証条件に近い濃い反応雰囲気 (バランスガスの分圧 0.2 以下) で予測 C₂ 収率が高い触媒組成を抽出した。すなわち、触媒評価条件は予測条件ではなく基準となる反応条件 (CH₄+O₂ 濃度 : 約 90%、バランスガス濃度 : 約 10%) を設定し、同条件下で NaMnW/SiO₂ 触媒が C₂ 収率 17.6% を発現することを指標として、実験実証による材料の良し悪しを評価した。その結果、36 触媒の実験実証の中で 8 触媒が同反応条件下で 15% 以上の C₂ 収率を発現したが、主に X-Mn-W/SiO₂ (X はアルカリ金属) の組成の触媒グループだった。次いで、実験検証により得られた 229 点のデータ (C₂ 収率 5% 以上のもの) を追加したデータセット(2)を新たな学習データとして SVR(2)を作成し、再度 C₂ 収率上位の触媒組成を予想させた。その結果、最大 C₂ 収率予想値は 21.73% へ僅かに空間が狭まり、18.00% 以上の C₂ 収率が期待できる触媒は僅か 3 触媒だった。ま

た、この3触媒はいずれも SVR(1)の時に既に実験実証している触媒組成だった。C2 予測収率の閾値を 16.00%以上へと広げて再度予測を行ったところ、91 触媒（うち 14 組成は SVR(1)の実証と重複）を新たに提案した。しかしこの時、3 元素担持触媒（M1-M2-M3/Support）の予測をターゲットとしているにも拘らず 50 触媒は 2 元素（M1-M2/Support）の組成となった。このことから SVR(1)と SVR(2)では、予測空間の減少、触媒組成における多様性の減少、実験実証における高収率触媒数の減少が認められる結果となった。

元データと追加データの傾向を図 4 に示す。HTS データと文献データから成るデータを前処理した 2,842 データは、データ数の 50%が C2 収率 8%以下、95%が C2 収率 20%以下から成るデータセットであり（図 4 A）、このデータセットを基に学習機 SVR(1)が作成されている。一方、学習機 SVR(2)の作成のために追加されたデータ分布では、データ数の 85%が C2 収率 14%以下を占めるデータであった（図 4 B）。従って、SVR(1)の実験検証の後に追加された実験データは C2 収率が高い領域（例えば 15%以上）を補完せず、C2 収率が 10%程度の領域により影響を与えるデータであるため、学習機 SVR(2)の段階でさらに高い C2 収率を発現できる触媒予測を行うことへの困難さが増してしまったと考えられた。実験研究者は検証結果を多角的に評価し、新しい視点に基づく実験の再設計が可能である。一方でデータ駆動型の触媒探索は、データの相関性に基づく材料探索を基本としている。従って、より上振れの目標領域に近い数値（外れ値）をどのように積極的に出現させられるのかが、データ駆動型の探索の向上に必要な技術であると考えられた。また、One-hot encoding による材料記述も空間を狭める要因であったと考察される。触媒組成を元素の組み合わせのみで射影できる表現手法は、多様なバイアスを含むデータを一元的に処理できる利点がある一方で、元素の混合比や調製手法などの空間拡張が反映できないため、おのずと探索空間が制限を受けてしまうことになる。固体触媒の材料情報をどのように記述できるのか、記述子探索の重要性^{19,20)}も示唆する結果となった。

これまでの検討事例からデータ駆動型の触媒探索ではデータのトレンドを紐解くツールとしての活用が期待できるものの、高い予測精度を備えた機械学習モデルの設計が必ずしも目標を実現可能な触媒提案に繋がるわけでは無いこと、更に機械学習の手法についても外挿の壁や記述子設計の技術的な課題があること、を具体的な検討事例を基に考察した。また、固体触媒材料を題材として扱う上では、固体触媒の機能性の解

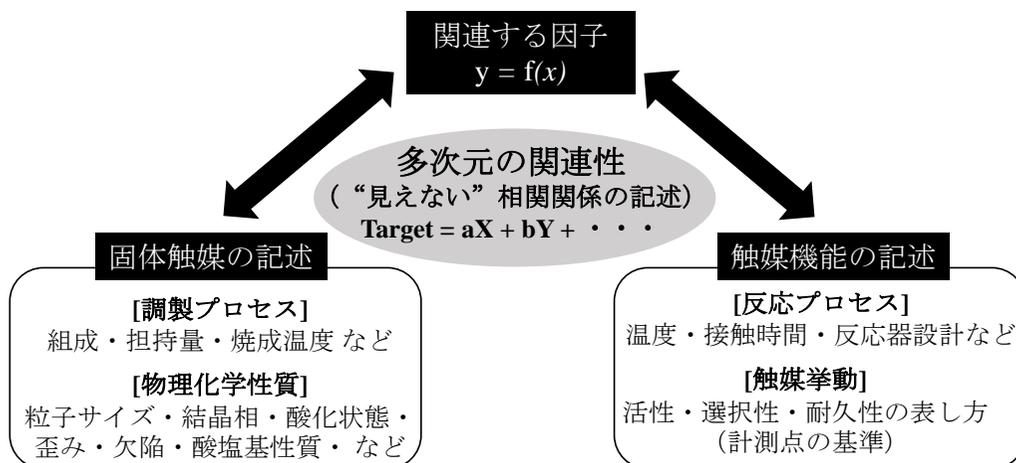


図5 固体触媒の機能表現（記述子設計）の多様さ

明が容易ではない点にも注意が必要である。図5に固体触媒の機能表現の多様性を例示した。データ駆動型の材料探索は、 $y = f(x)$ の相関関係を多次元で認識することで人知では見えにくい相関性や未探索空間での可能性を指南する手法である。X の説明変数（記述子）の妥当性の担保が y の目的変数を実現できる要である。しかし、固体触媒の振る舞いは一様には記述できず、もし妥当な共通の記述子（global descriptor）が発見されたとしてもそのデータ収集コスト、さらには合成実現性（逆問題）が新たな課題として生じる。従って、完全な過不足の無い相関関係の描写方法の確立に注力するよりも、「完全ではない機械学習ツールによる固体触媒の予測支援をどのように実装可能であるか？」を問うことがより重要（現実的）である。この問いに対するいくつかの解法の中で、『実験研究者の観点』を活かした実践について最後に紹介する。

OCM 触媒開発の異なるターゲットとして低温化 OCM 触媒の開発が挙げられる。中でも La_2O_3 系の OCM 触媒は $420\sim 600^\circ\text{C}$ 付近の低温から C2 収率の立ち上がりが見られることを特徴としている（ex. $\text{NaMnW}/\text{SiO}_2$ 系は $700\text{-}900^\circ\text{C}$ 付近での C2 生成能を有する）。しかし、これまでの La_2O_3 系の低温 OCM 触媒材料の探索範囲は狭く、広大な未探索空間での新たな発見の可能性がある²¹⁾。そこで、3 元素担持 La_2O_3 触媒（M1-M2-M3/ La_2O_3 ）を効率的に探索することで低温化 OCM 触媒の可能性を広げられないかと考えた^{22,23)}。機械学習の基礎データには先述の HTS 実験データ（40,339 点）^{14,15)} を用いた。しかし、これらの HTS データは高温 OCM 触媒探索を指向したデータ収集を展望したため $700\text{-}1000^\circ\text{C}$ でのデータ範囲により構成されている。従って、そのままでは

直接的には高温 OCM 触媒の予測指南にしか使用することが出来ない。さらに、ランダムな探索空間を設定した HTS 実験データ¹⁵⁾の未探索組成の組み合わせを内挿補完により提示する懸念もあった。この課題を解消するために、機械学習が元データの傾向を捉えやすい性質に着目し、高温 OCM 触媒に関する HTS 実験データを基に作成した学習機 (SVR) からの『知識抽出』を試行した。

「機械学習の手法はデータの傾向を捉えることに秀でている」との観点から、高温 OCM 触媒データを学習し、C2 収率が (高温で) 高いと予測された M1-M2-M3/La₂O₃ 触媒について、「OCM 反応雰囲気での La₂O₃ の特徴を引き出しやすい (相性が良い) M1-M2-M3 を捉えている」との仮説を立てた。One-hot encoding による触媒記述方法 (図 3) を基に実際に SVR 空間での予測 C2 収率 (700-1000°C の温度域) が高いものから 20 触媒を実験実証した結果、高温域での予測 C2 収率の大小と低温域での立ち上がり温度の出現頻度 (すなわち La₂O₃ の低温 OCM 特性の促進) に必ずしも相関性は認められなかったが、金属を担持していない none/La₂O₃ 触媒の C2 立ち上がり温度 500°C に対して、同反応条件下で 450°C から C2 収率を発現する組み合わせが 11 触媒確認できた (La₂O₃ の低温 OCM 動作能を促進する M1-M2-M3 組成の発見)。更に、材料探索空間の拡張性を補う方策として文献データ (1,868 点)⁹⁾の追加を提案した。前述のように One-hot encoding の課題の一つは、0/1 の記述のための index ラベルで制限された材料探索空間に縛られる点である。図 2 に例示したように、HTS 実験データで定義された限られた材料探索空間を文献データのより広い探索空間によりラベル拡張することで、より広い探索空間での予測に先導できると発想した。拡張後のデータセットを基にした SVR 学習機では 12 触媒の実証で 2 触媒しか低温 La₂O₃ 触媒の機能を示さなかったが、HTS データには含まれない新しい構成元素 (Yb や Ga) が含まれる触媒を発見できた。また、異なる手法の学習機 (RFR) でも検討した結果、12 触媒中 11 触媒が低温 La₂O₃ 触媒の性質を示した (但し、この場合は GaHf に偏った触媒組成が頻出した)。以上の結果、知識抽出の視点を活かした低温 La₂O₃ 触媒の探索を行うことで、低温駆動能力を促進できる 24 種類の M1-M2-M3/La₂O₃ 触媒の発見に至ることができた。

データ駆動型の触媒開発手法は、これまでの因果関係の解明を主とする材料探索アプローチに新たな視点をもたらすツールとして活用が期待される。機械学習・データマネジメントの得意とする分類、回帰、パターン解析、可視化などの手法を展開することで、触媒化学分野に新しい知の創出をもたらすきっかけとなることを期待したい。

謝辞

戦略的創造研究に推進事業チーム型研究 (CREST)・「実験・計算・データ科学の統合によるメタン変換触媒の探索・発見と反応機構の解明・制御」(No. JPMJCR17P2、H29.10-R6.3) の支援の下、高橋啓介教授 (北海道大学、研究代表者)、大山順也准教授 (熊本大学)、谷池俊明教授 (北陸先端科学技術大学院大学)、宇野毅明教授 (情報学研究所) との活動の中で議論を深め、検討内容を精錬することで研究成果として報告できたものである。ここに謝意を表す。

引用文献・参考文献」等

- [1] K. Takahashi, L. Takahashi, I. Miyazato, J. Fujima, Y. Tanaka, T. Uno, H. Satoh, K. Ohno, M. Nishida, K. Hirai, J. Ohyama, T. N. Nguyen, S. Nishimura, T. Taniike, *ChemCatChem* 2019, 11, 1146.
- [2] T. Toyao, Z. Maeno, S. Takakusagi, T. Kamachi, I. Takigawa, K. Shimizu, *ACS Catal.* 2020, 10, 2260.
- [3] B. MacQueen, R. Jayarathna, J. Lauterbach, *Curr. Opin. Chem. Eng.* 2022, 36, 100781.
- [4] W. Beker, R. Roszak, A. Wołos, N. H. Angello, V. Rathore, M. D. Burkner, B. A. Grzybowski, *J. Am. Chem. Soc.* 2022, 144, 4819.
- [5] K. Takahashi, J. Ohyama, S. Nishimura, J. Fujima, L. Takahashi, T. Uno, T. Taniike, *Chem. Commun.* 2023, 59, 2222.
- [6] G. E. Keller, M. M. Bhasin, *J. Catal.* 1982, 73, 9.
- [7] W. Hinsen, M. Baerns, *Chem. Ztg.* 1983, 107, 223.
- [8] E. V. Kondratenko, T. Peppel, D. Seeburg, V. A. Kondratenko, N. Kalevaru, A. Martina, S. Wohlrab, *Catal. Sci. Technol.* 2017, 7, 366.
- [9] U. Zavyalova, M. Holena, R. Schlogl, M. Baerns, *ChemCatChem* 2011, 3, 1935.
- [10] K. Takahashi, I. Miyazato, S. Nishimura, J. Ohyama, *ChemCatChem* 2018, 10, 3223.
- [11] S. Nishimura, J. Ohyama, T. Kinoshita, S. D. Le, K. Takahashi, *ChemCatChem* 2020, 12, 5888.
- [12] J. Fujima, Y. Tanaka, I. Miyazato, L. Takahashi, K. Takahashi, *React. Chem. Eng.* 2020, 5, 903.
- [13] S. Nishimura, X. Li, J. Ohyama, K. Takahashi, *Catal. Sci. Technol.*, 2023, 13, 4646.

- [14] T. N. Nguyen, T. T. P. Nhat, K. Takimoto, A. Thakur, S. Nishimura, J. Ohyama, I. Miyazato, L. Takahashi, J. Fujima, K. Takahashi, T. Taniike, *ACS Catal.* 2020, 10, 921.
- [15] T. N. Nguyen, S. Nakanowatari, T. P. N. Tran, A. Thakur, L. Takahashi, K. Takahashi, T. Taniike, *ACS Catal.* 2021, 11, 1797.
- [16] S. Mine, M. Takao, T. Yamaguchi, T. Toyao, Z. Maeno, S. M. A. H. Siddiki, S. Takakusagi, K. Shimizu, I. Takigawa, *ChemCatChem* 2021, 13, 3545.
- [17] J. Ohyama, S. Nishimura, K. Takahashi, *ChemCatChem*, 2019, 11, 4307.
- [18] A. Cruellas, T. Melchiori, F. Gallucci, M. van S. Annala, *Catal. Rev.* 2017, 59, 234.
- [19] S. Ishioka, A. Fujiwara, S. Nakanowatari, L. Takahashi, T. Taniike, K. Takahashi, *ACS Catal.* 2022, 12, 11541.
- [20] F. Garcia-Esvobar, T. Taniike, K. Takahashi, *J. Chem. Inf. Model.* 2024, 64, 1512.
- [21] J. Ohyama, T. Kinoshita, E. Funada, H. Yoshida, M. Machida, S. Nishimura, T. Uno, J. Fujima, I. Miyazato, L. Takahashi, K. Takahashi, *Catal. Sci. Technol.* 2021, 11, 524.
- [22] S. Nishimura, S. D. Le, I. Miyazato, J. Fujima, T. Taniike, J. Ohyama, K. Takahashi, *Catal. Sci. Technol.* 2022, 12, 2766.
- [23] S. Nishimura, High-Throughput Screening and Literature Data Driven Machine Learning Assistaing Discovery of La₂O₃-based Catalysts for Low-Temperature Oxidative Coupling of Methane, Proceedings of the 31st Annual Saudi–Japan Symposium on Technology in Fuels & Petrochemicals –Innovative Catalyst Development, KFUPM, Dhahran, Saudi Arabia, 2022, pp. 32–42.