

計算社会科学を基盤とした政策科学の新展開



東京大学 工学系研究科教授・総長特別参与 坂田一郎

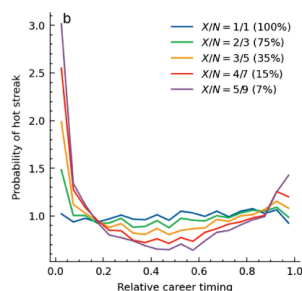
自己紹介: コンピューテーション & 社会科学

《 研究内容 》

- 深層学習、自然言語処理、複雑系ネットワーク科学等の**データドリブンなアプローチ**を組み合わせ用い、人の判断とそれに基づく行動、人や組織の行動の間のインタラクション、それらの集まりとしての社会現象を解析し、**再現可能なパターンの発見やそれを用いた予測**に取り組む。また、開拓した知見や技法を活用して**実用的な意思決定支援ツール**を開発。
- 科学者や科学コミュニティは、データが最も構造化され、整備された領域。それらを対象とした研究として、「ホットストリーク」、「成果の認知の遅れ」、「スタートアップ・レディネス」、「研究トピックの先進性」、「技術融合の特定」などのテーマで出版。世界で「**サイエンス オブ サイエンス**」と呼ばれる領域が発展中。[D. Wang, A.L. Barabasi, 『サイエンス オブ サイエンス』(森北出版、2025)]

《 専門領域 》

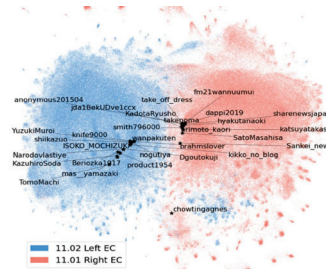
科学計量学



— Hot Streak —

Sci. Rep. 14(1), 2024, 28172

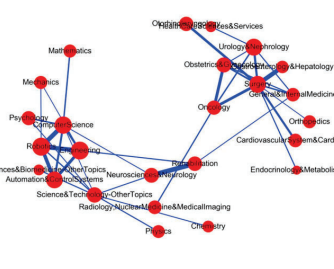
計算社会科学



— Echo Chamber —

Sci. Rep. 11(1), 2021, 7491

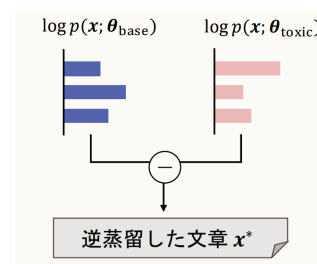
技術経営



— Tech Convergence —

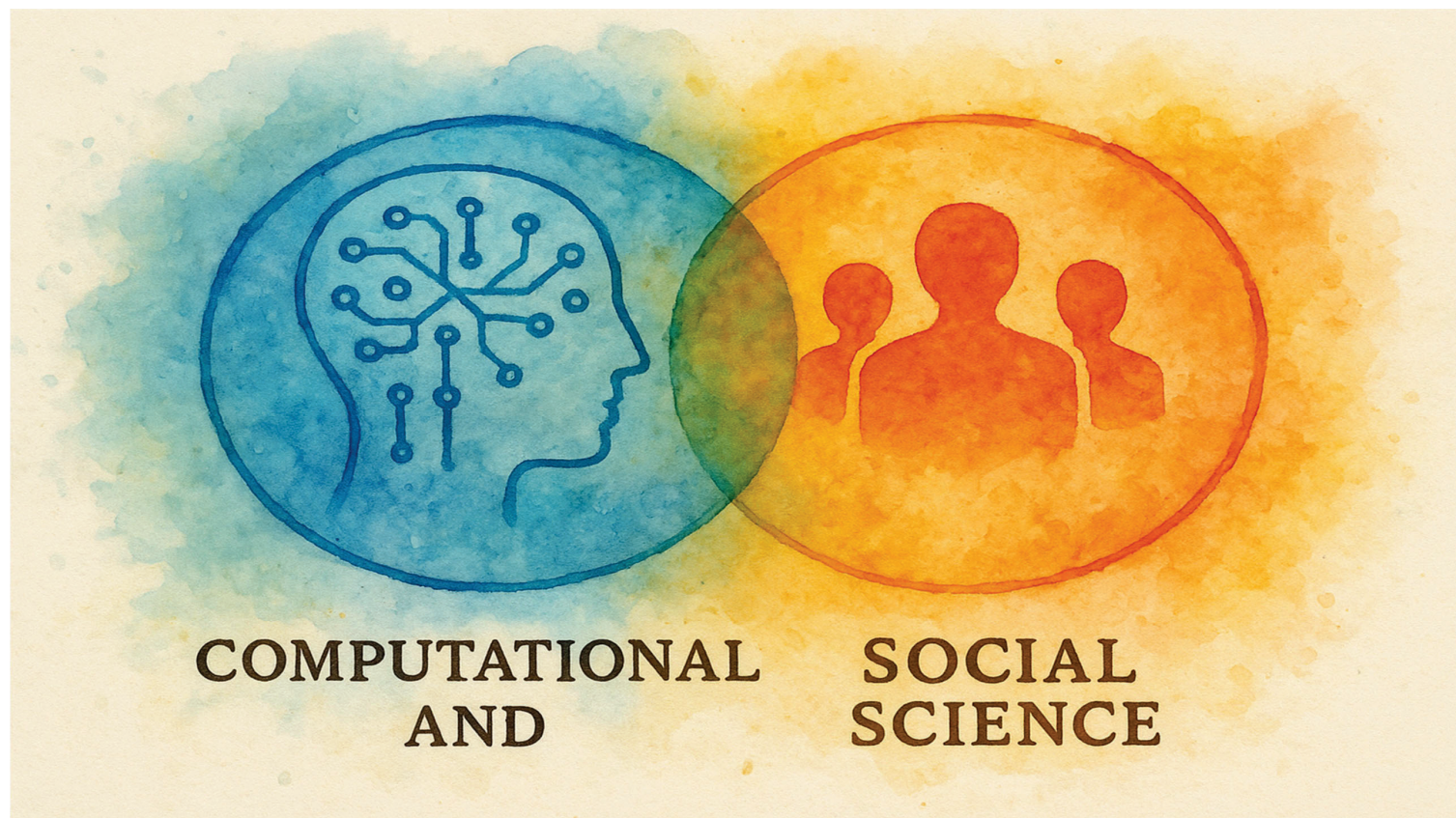
TFSC 146, 2019, 751-766

自然言語処理/LLM



— AI & Ethics/ Detoxification —

ICLR 2025
(深層学習のTOPカンファレンス)



計算社会科学とは

- 「計算社会科学」は、データ分析・モデリング・シミュレーションなどの**データドリブンなアプローチ**で、人の社会における行動、人や組織の行動の間のインタラクション、それらの集まりとしての社会現象を解析し、再現可能なパターンの発見や予測をすることを目的としている。
- デジタル革命の進展とともに、**計算科学と社会科学の融合**により成立。2009年のScience誌上で発表された著名な論考”Computational Social Science”が領域の名称を定めた。計算社会科学により、人や、コミュニティ、組織、社会を対象として、それまでになった広さと深さで、データを集積し、分析することを可能に。
- 価値のあるパターンの発見や予測には、対象とするドメインにおける**社会的・文化的な文脈や慣行**を理解することが不可欠。それが欠けた情報は、C.マスビアウが言う「薄いデータ」に過ぎなくなる。



—事例：国・地域ごとに異なる炭素排出の文脈—

例) アフリカ:調理やバイオマス、アジア:人口と消費拡大、北米:輸送、欧州:持続可能性や幸福



(Source) Y. Long, L. Huang, S. Montaga, Z. Chen, X. Ding, Y. Shigetomi, T. Zhao, K. Asatani, I. Sakata, Y. Yoshida, "Tracing the evolution of household carbon emission research by machine learning", **Environmental Impact Assessment Review** 115 (2025), 108039.

計算社会科学の対象となる代表的データ

SNS内の会話



Amazonレビュー文



科学技術文献



社会ネットワークと交流



人々の移動・交通



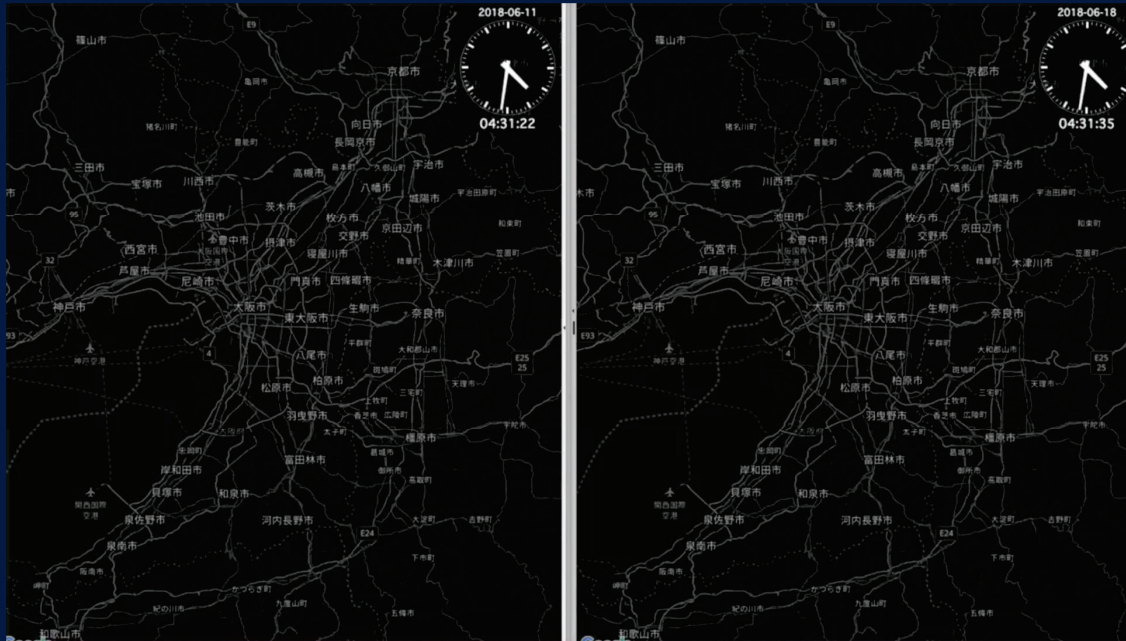
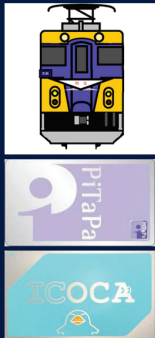
イノベーション活動



実データ例:鉄道移動データによる大阪北部地震の観測(2018.6.18)

平常時

地震発生(AM 7:58)

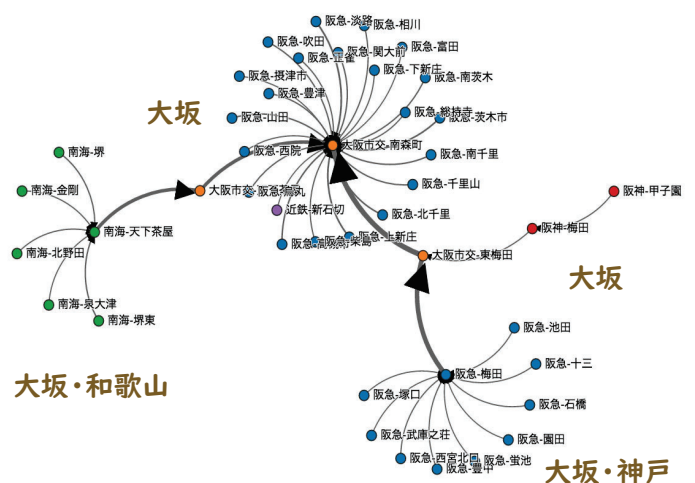
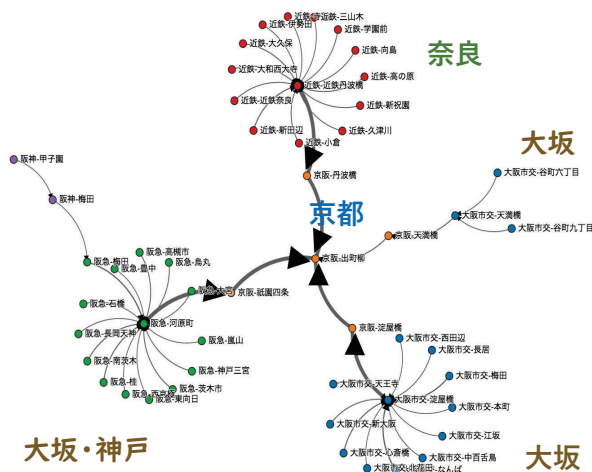


(source) T. Maeda, J. Mori, M. Ochi, I. Hayashi, T. Sakimoto and I. Sakata "Comparative Examination of Network Clustering Methods for Extracting Community Structures of a City from Public Transportation Smart Card Data", *IEEE Access* vol.7 (1) (2019) , 53377-53391.

大坂の人々は京都の祭りに行くが、その逆は少ない

五山の送り火(京都)

天神祭(大阪)



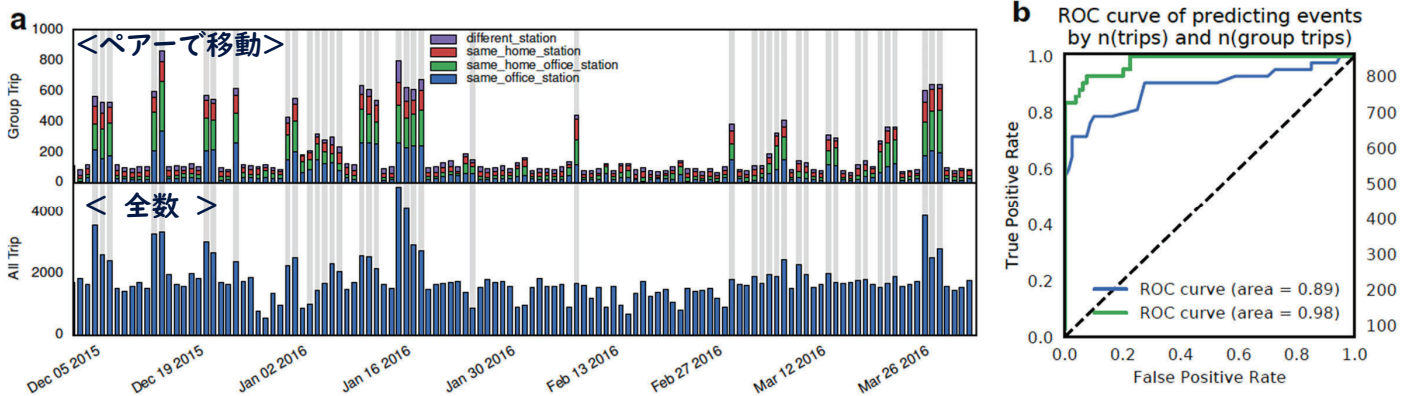
なぜ違いが生まれるのか?の理解には、文化的文脈の知識が必要

「知り合い」？ が分かったら価値が生まれる

ある駅でイベントが行われたかどうかは、駅の利用の総数ではなく、**ペア利用（知り合いとの移動）の数**だけを用いて観察した方が、高い精度で予測できる（右図、**緑色ライン**）。

それでは、ペア移動をどうやって特定するのか？ アイデアの出どころ。

質問：60代の男性は、同世代の女性と比較して、同世代の男性（友人）、10代・20代の男女（子供や孫）との移動の頻度は少ないです。では、だれが一緒（ペアで）に電車に乗ってくれているのでしょうか？



(Source) K. Asatani et al. "Detecting Interpersonal Relationships in Large-Scale Railway Trip Data", *Journal of Computational Social Science* 1 (2) (2018), 313-326.

計算社会科学の対象となる代表的手法

社会ネットワーク分析



オンライン上の自然実験の分析



書誌と属性情報からの知見抽出



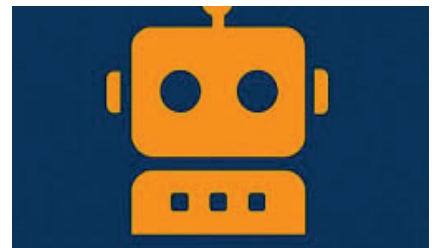
社会現象のシミュレーション



予測モデルの作成

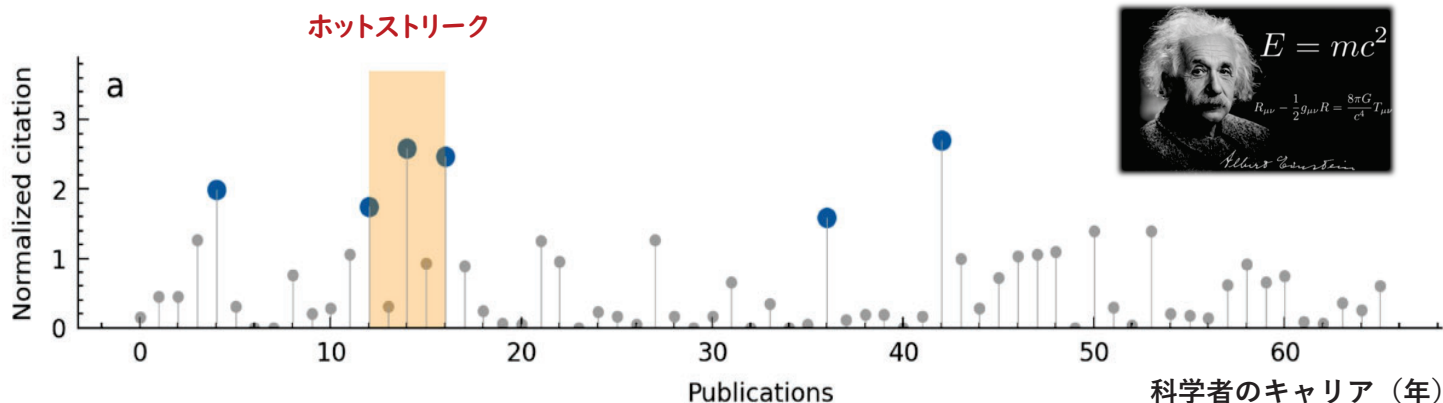


LLMエージェントを用いた仮想実験



科学者のホットストリークの時期は？

人には、連続して成功が訪れる「調子の良い時期 (Hot Streak)」が存在している。
 スポーツや芸術などと並んで、科学者にもそれがあると考えられている。
 有名なのは、アインシュタインが彼のキャリアを代表する4つの画期的論文を発表した1905年。
 1回あたりの平均の長さは平均3.7年、約9割の科学者がキャリアの中で一度は経験する。

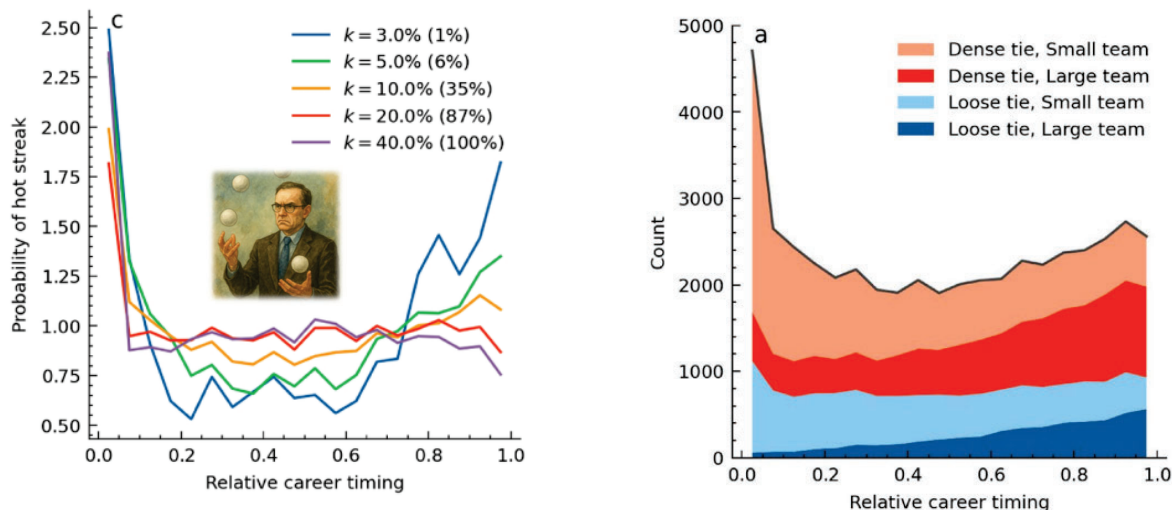


(source) L. Wang et al., "Hot Streaks in artistic cultural, and scientific careers", *Nature* 559(7714) (2018), 396-399.
 N. Hlgashide, T. Miura, Y. Tomokiyo, K. Asatani, I. Sakata, "Mid-career pitfall of consecutive success in science", *Scientific Reports* 14 (2024), 28172.
 東京大学統合報告書 2023, 78-79.

11

キャリアの初期と最後に多いホットストリーク

我々は、全Scopusデータを用いた分析から、「ホットストリーク」がキャリアの**初期**と**最後**に多いことを発見。
 特に、初期は、分野を分けて分析してみても、大層の分野で確率が高いことがわかる。
 初期には**濃密な人間関係**、後期には**大きなチーム**を率いていることがそれに寄与していると考えられる。
 また、ホットストリーク事例では、一貫して研究トピック数は、そうでない場合と比べて少ない。



(source) N. Hlgashide, T. Miura, Y. Tomokiyo, K. Asatani, I. Sakata, "Mid-career pitfall of consecutive success in science", *Scientific Reports* 14 (2024), 28172.

12

【事例2 科学論文DB×つながり×計算科学】

国別の研究トピックの先進性の計測

目的: 論文の量だけで、世界的にみて研究をリードしているかどうかは判断できない。そこで、国別に研究トピックの「**先進性**」を**定量化**することで、その判断材料を提供する。

データ: 2021年までの全分野の論文とその参考文献リスト全体

手法: 論文群の参考文献リストのトピックを比較、その**類似性**を計算

国別、年別に、出版論文の参考文献リストを積み上げ

1論文あたりの参考文献数で重み付け

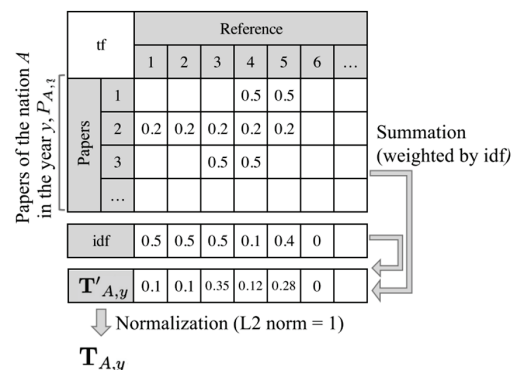
さらに、TF-IDFを用いて被引用数が高い参考文献に低い重み付け

被引用数1000件以上の参考文献は除いて計算

L2 norm = 1に正規化

⇒ 国Aのy年の出版論文の参考文献リスト: $T_{A,y}$

a Calculate research topic $T_{A,y}$ of nation A in year y.

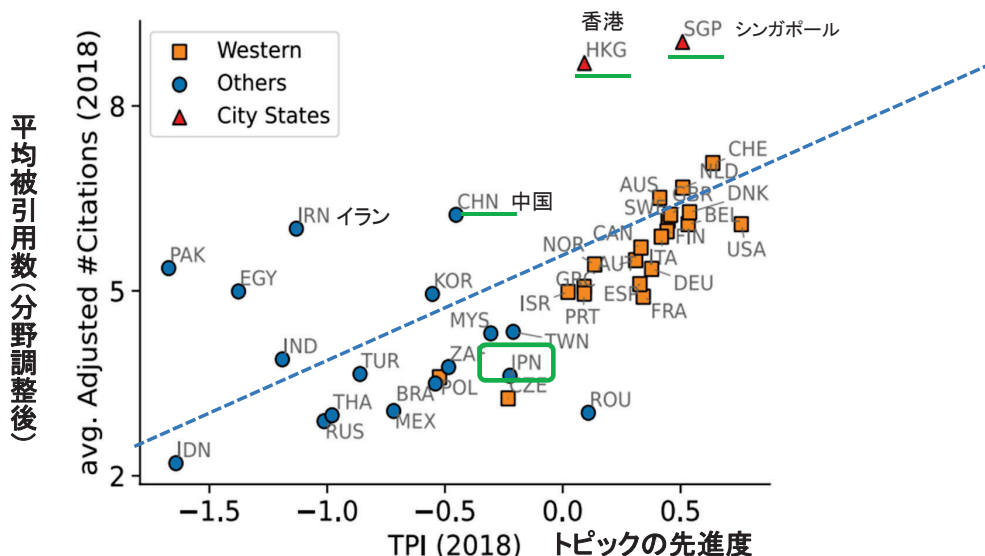


- トピックの類似性をもとに、他国すべてを対象としたトピックの先進度を示す**独自指標(TPI)**を計算

(source) K. Asatani, S. Oki, T. Momma, I. Sakata, "Quantifying progress in research topics across nations", *Scientific Reports* 13 (2023),4759.

「研究トピック」の先進性と平均引用度は相関する

各国のトピックの先進度 (TPI) と論文の平均被引用数との間には、一定の相関がある。シンガポール、香港、中国は、平均引用数が上方に乖離する傾向が見られる。

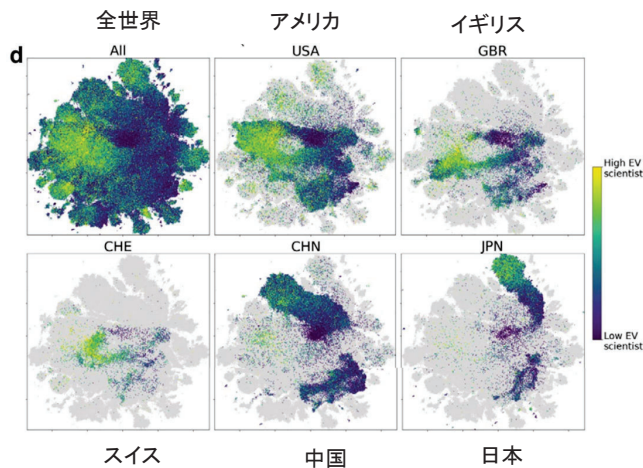


(source) K. Asatani, S. Oki, T. Momma, I. Sakata, "Quantifying progress in research topics across nations", *Scientific Reports* 13 (2023),4759.

インフルエンシャルな研究者が多いとトピックは先進する

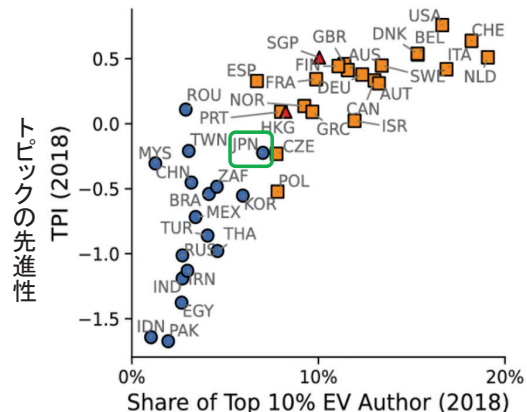
国別にみると、自国内での**インフルエンシャルな研究者**（共著ネットワーク内でEV中心性の高い研究者）の比率と**TPI（トピックの先進度）**との間には、高い相関が観察される。

研究トピックを主導するにあたり、「**頭脳循環**」の拡大や学術活動の**国際化**の重要性を示唆する結果。



世界の研究者ネットワークと各国の存在感

(source) K. Asatani, S. Oki, T. Momma, I. Sakata, "Quantifying progress in research topics across nations", *Scientific Reports* 13 (2023), 4759.



インフルエンシャルな研究者の密度

(備考) 横軸をTOP 1%、20%のシェアとしても同様な結果が得られる。

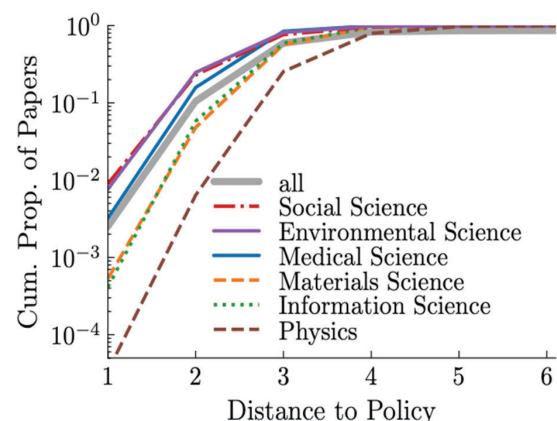
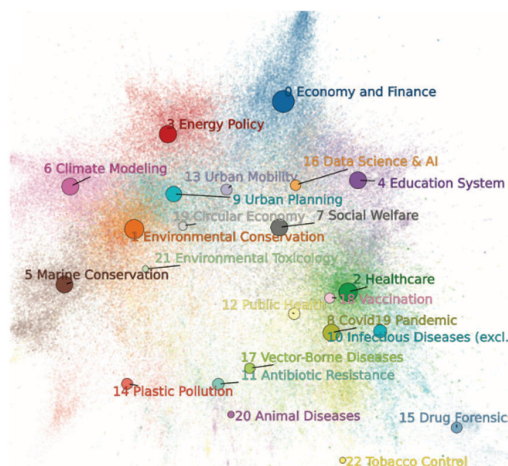
【事例3 2DB統合×政策/ルール×計算科学】

IGO政策文書で引用される学術知識

学術論文と**政策文書**の大規模データ（2015-2023）を統合。国際機関（IGO）の政策文書群に引用されている23万件の論文群を引用ネットワーク分析によりクラスターへと分割（専門領域に分類）。

自動分類されたクラスターは、**医療・ヘルス**、**環境**、**社会**に大別できる（左図）。

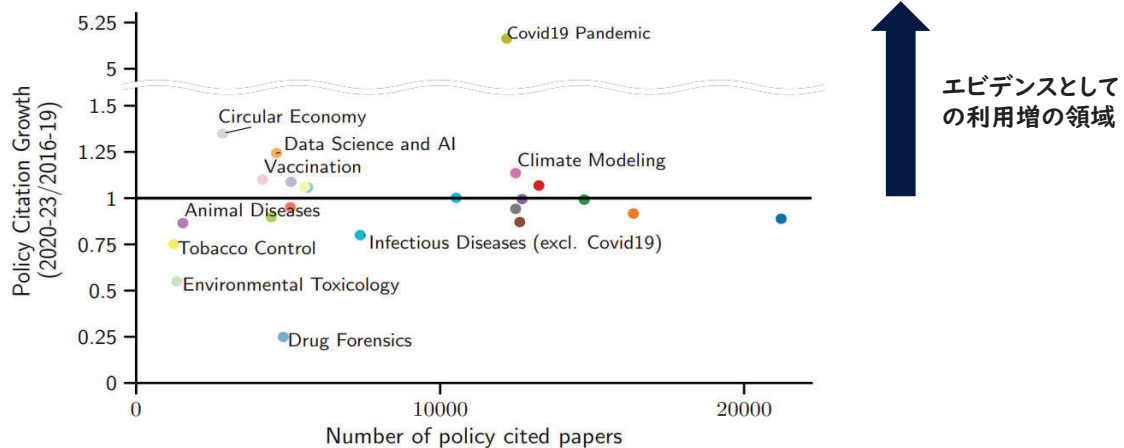
分野により、政策との距離は、かなり異なる。近いのは「**環境科学**」、遠いのは「**物理学**」（右図）。



(source) K. Asatani, Y. Iwata, Y. Tomokiyo, B. Mahfouz, M. Yarime, I. Sakata, "Influential scientists shape knowledge flows between science and IGO Policy". <https://arxiv.org/abs/2506.06753>

国際機関が求める科学的ニーズへの対応とそのインパクト

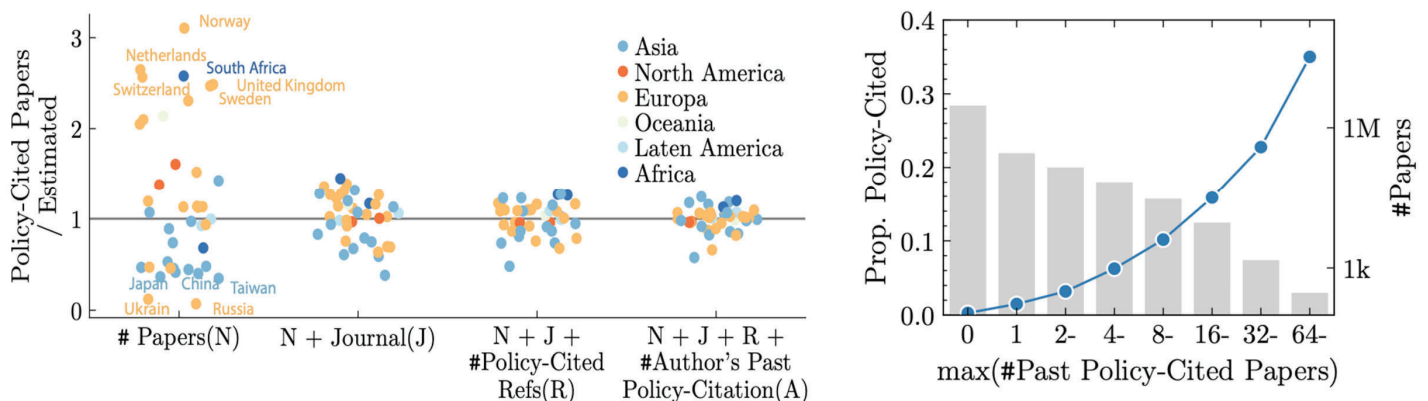
2020年以降、Covid-19関係の科学的知見が非常に多く、**エビデンス**として利用されている。
Circular Economy、Data Science・AI、Climate Modeling等の引用が伸びておりニーズの拡大がわかる。
成果の公共的利用はインパクトを高める傾向※1、Covid-19領域では科学と政策のCoevolutionを観測※2
課題解決への貢献が学術のインパクトを高める。**アクセラレーション活動**の価値を示す。



(source) K. Asatani, Y. Iwata, Y. Tomokiyo, B. Mahfouz, M. Yarime, I. Sakata, "Influential scientists shape knowledge flows between science and IGO Policy". <https://arxiv.org/abs/2506.06753>
 ※1 : Y. Yin et al., "Public use and public funding of science", *Nature Human Behavior* vol.6 (2022), 1344-1350.
 ※2 : Y. Yin, J. Gao, B. F. Jones, D. Wang, "Coevolution of policy and science during the pandemic", *Science* vol.371 (issue 6525) (2021), 128-130.

政策文書での被引用数は4つの要素で説明できる

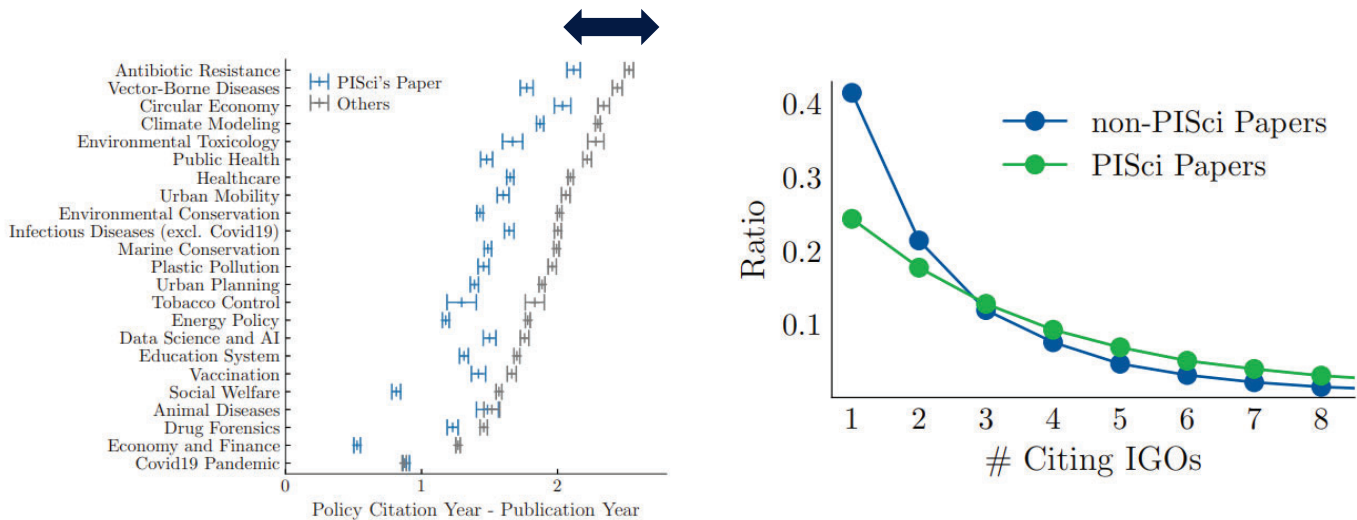
政策文書での論文引用数は、国別には、論文数 (N)、掲載ジャーナル (J)、政策引用された参考文献 (R)、著者の過去の政策文書での被引用 (A) の**4つの要素**で、かなり良く説明できる。
過去に64回以上、被引用されている著者を含む論文は、次の論文は40%近い確率で引用される。
グローバルサウスの研究成果は、直接的には届きにくい。



(source) K. Asatani, Y. Iwata, Y. Tomokiyo, B. Mahfouz, M. Yarime, I. Sakata, "Influential scientists shape knowledge flows between science and IGO Policy". <https://arxiv.org/abs/2506.06753>

インフルエンシャルな科学者の研究成果は早く、広く伝わる

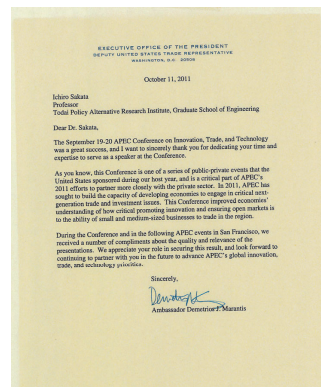
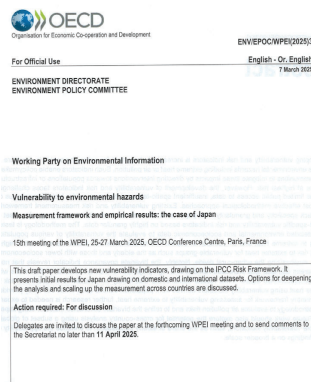
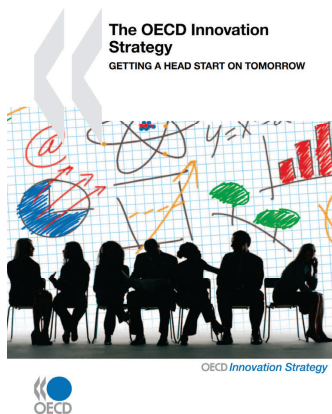
特定の科学者の論文群に、政策文書での引用は偏っている。例えば、Climate Modelingでは、わずか0.62%の科学者で被引用30%を占めている。これらを**Policy-Influential Scientists (PI-Sci)**と定義。PISciのペーパーは、**早く(左図)、広く(右図) 引用される傾向**が鮮明。



(source) K. Asatani, Y. Iwata, Y. Tomokiyo, B. Mahfouz, M. Yarime, I. Sakata, "Influential scientists shape knowledge flows between science and IGO Policy". <https://arxiv.org/abs/2506.06753>

自身のグローバルな政策形成への関与の経験

IGOの政策形成活動に対する「手触り感」は、研究を進める上で役に立つ
分析の企画や分析結果を活用する上でも有益。産官学協働の経験や回転ドアの重要性を示す。



Expert Advisory Group Member 1)
(2010)

OECD共同研究
(2024-2025)

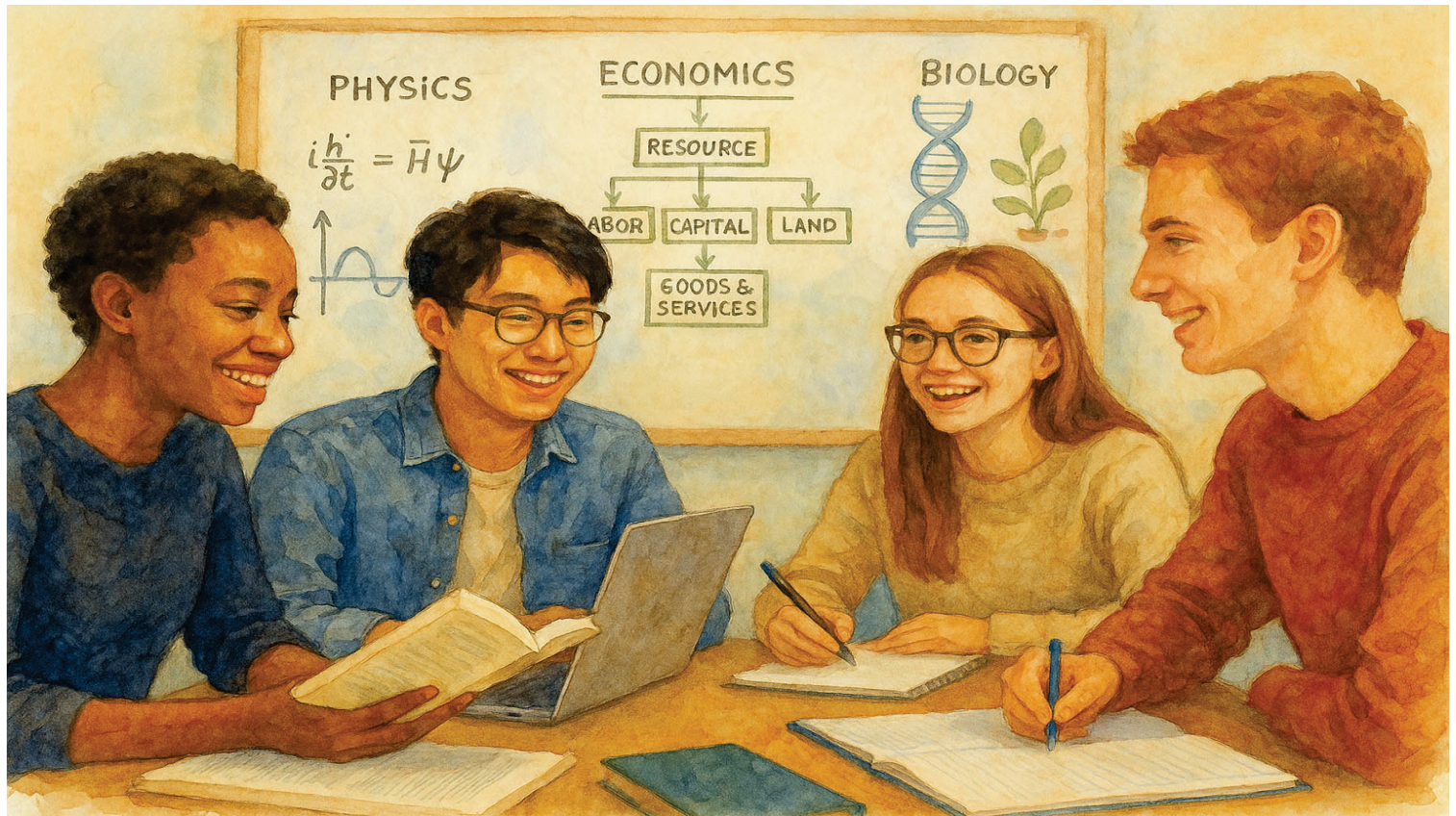
APEC準備会合でのインプット 2)
(2011)

政策提言の論考 3)
(2016)

1) OECD HP: https://www.oecd.org/en/publications/the-oecd-innovation-strategy_9789264083479-en.html

2) 国際枠組み Asia-Pacific Economic Cooperation

3) M. Sugiyama, I. Sakata, H. Shiroyama, H. Yoshikawa and T. Taniguchi, "Five years on from Fukushima", *Nature* 531(2016), 29-31.



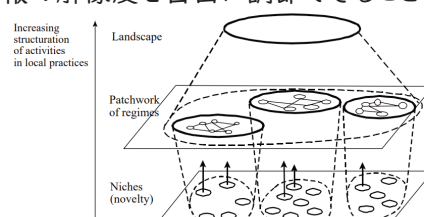
テクノロジー・インフォマティックスの展開

○科学技術政策又は技術経営の意思決定支援に用いられてきた「科学計量学」や「サイエンス・オブ・サイエンス」は、主として、科学技術主導のイノベーションを捉えるための手段である。
従って、現代の企業社会や政府が直面しているイノベーションの**3つのDriven**（テックに加え、ミッション、シティズン）を総合的に捉えようとすると、その範囲の拡張が必要となる。

○そこで我々は、幅広く社会の情報を扱う「**計算社会科学**」の枠組みを基礎・基盤として、その応用領域となる「**テクノロジー・インフォマティックス**」を提案している。

インプットとなる情報としては、①科学技術ビックデータ、②経済・社会の変化に関する情報、③制度や規制の関連情報の3種類を視野に置き、イノベーションを左右する**技術の重要な進歩、経済・社会の潮流、社会システムの動向**の3側面をカバーする。※ F.W. GeelsのMLPのフレームワーク(F. W. Geels, TFSC 72(2005), pp.684、下記右図)に対応

○**3つの側面**に関する**将来予測**、社会や技術に生じている**アノマリー**や**トレンド変化の早期把握**、変化において**鍵となる組織や個人**の特定、**急成長や融合進展の技術領域**、**統合により価値が高まるパートナー**などの知見の抽出・提供を通じ、意思決定支援を行う。顕微鏡のように、情報の解像度を自由に調節できることが重要。



(Source) F.W. Geels, "Processes and patterns in transitions and system innovations : Refining the co-evolutionary multi-level perspective", TFSC 72 (2005), 681-696.

計算社会科学からテクノロジー・インフォマティクスへ

計算社会科学の諸手法や汎用的知見、ノウハウ

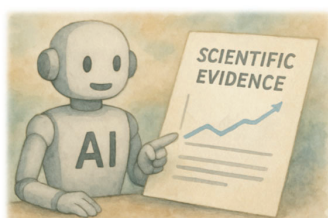
大規模な科学技術文献の分析手法



政策に関する科学者ネット及び
インフルエンサーの理解手法



最新のAI手法研究



M&Aに関する推薦手法



- ・技法やノウハウの展開
- ・ドメイン専門知との融合
- ・知の俯瞰、予測、特定

■政策や技術経営の意思
決定支援ツールへと実装
■ドメイン知識と統合



テクノロジー・インフォマティクスと意思決定支援

インプット情報

科学技術情報

・学術論文DB・特許DB・2次DB



経済・社会情報

・SNS情報、ニュース、企業DB
・SDGs関連論文、技術系起業等



政策/規制情報

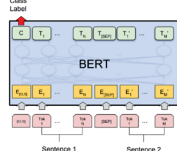
・国際政策文書DB・規制関連論文群
・標準、ガイドランなど



情報の構造化・統合

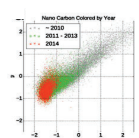


分析・予測・知見抽出



- ・自然言語処理
- ・ネットワーク解析
- ・生成モデル

ニーズに応じた提示・橋渡し



戦略的インテリジェンス の提供、意志決定支援

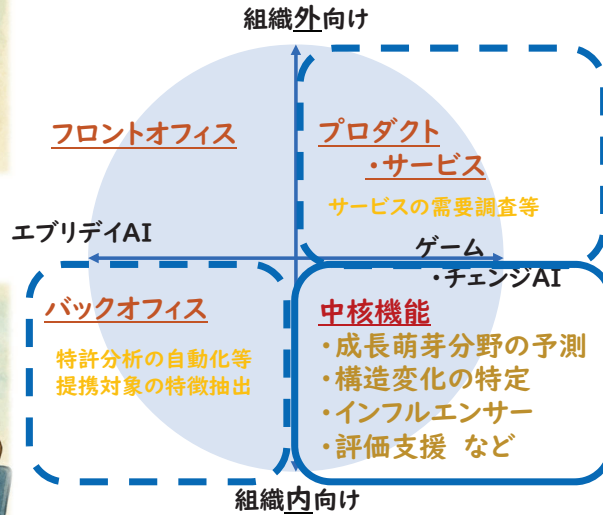
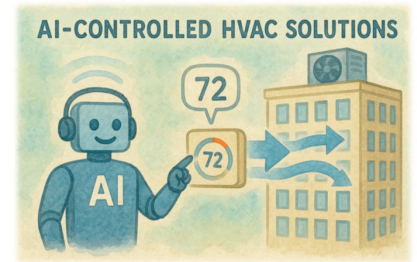
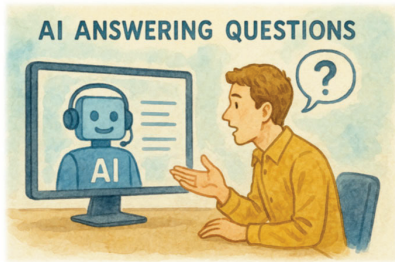
意思決定のフロー



意思決定プロセスの
適切な「場」への埋め込み
「LLM (AI) エージェント」が効果的にアシスト
人間による業務の代替と拡張の双方

テクノロジー・インフォマティックスの主対象領域

ガートナー社
AIオポチュニティ・ラダー



(Source) <http://gartner.co.jp/ja/topics/ai-readiness> に加筆

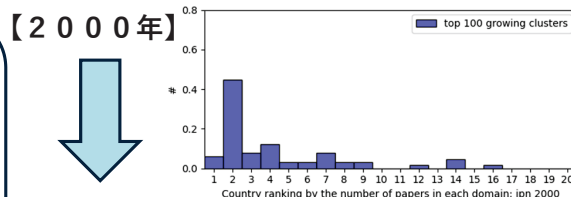
【TI事例①:データ取り扱いのノウハウの展開】

成長トップ100科学領域における日本の存在感

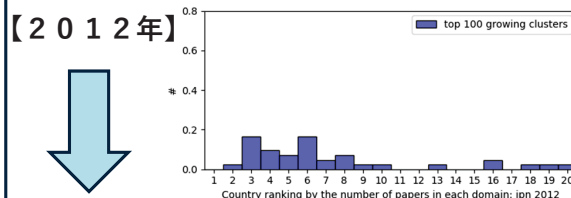
・Scopus (2024年度は8700万論文、10億件の引用)に基づき内容的な近さに応じ約10,000領域に分類(※再帰的にクラスタリングし、論文数4万件以下のクラスタに分類)

・成長トップ100科学領域 = 「当該年度の論文数 / 直近10年の論文数」が多いトップ100クラスタ (対象は、当該年度の論文数>1000報に限る)

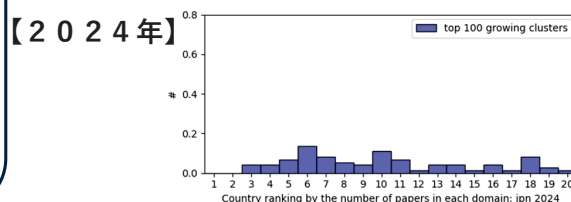
・クラスタ名はGPT-4oを用いて付与



1位: Fluidized Bed and Particle Dynamics, Nuclear Fission Physics, High-Temperature Superconductivity in YBaCuO, Superfluid Helium and Thermodynamics
2位: Optical Fiber Sensing and Polarization, Detector and Sensing Technology, Mitochondrial ATP and Enzyme Activities, Phase Change & Heat Transfer, Antibiotic Resistance in Staphylococci...



2位: Photocatalytic Surfaces and Materials, Targeted Therapies for Lung Cancer
3位: Embryonic Stem Cell Research, Cancer Stem Cell Research, Visible Light Photocatalysis, Human Motion and Gesture Recognition
...
下位: Gender and Sports Studies, Leadership and Team Performance, Biodiesel Production and Emissions, Children's Physical Activity and Fitness

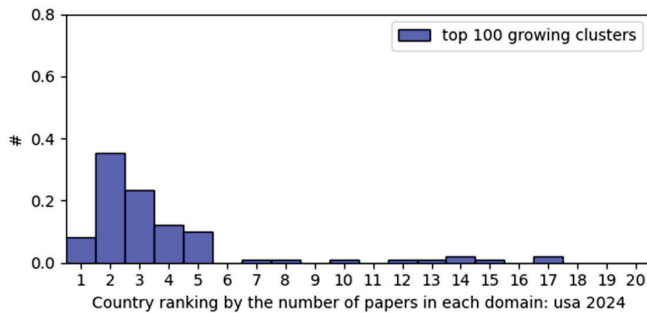


3位: Machine Learning for Materials Science, Programmed Cell Death and Ferroptosis, Genetic Prognostic Indicators in Cancer
4位: Colorectal Disease Diagnosis in Endoscopy, Neural Networks and Nonlinear Dynamics, Non-Alcoholic Fatty Liver Disease
...
下位: AI in Education, Digital Twins in Manufacturing, Protein Interaction Prediction, Plastic Pyrolysis & Recycling, Biodiversity and Climate Change

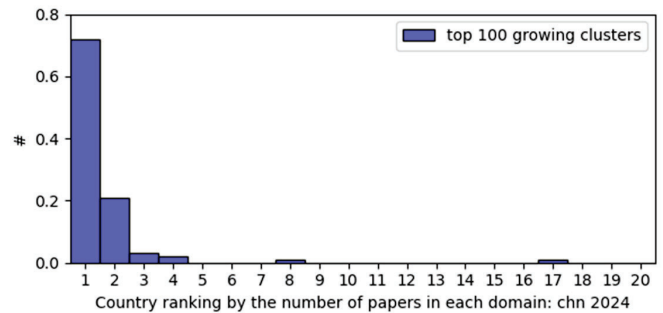
(出典) 坂田・浅谷・西本 研 浅谷公威特任准教授作成 科学技術・学術審議会 研究費部会(2025.7.28)資料

最新時点（2024年）での4カ国比較

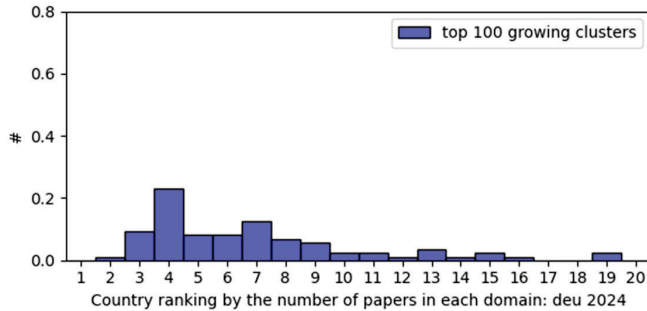
<アメリカ>



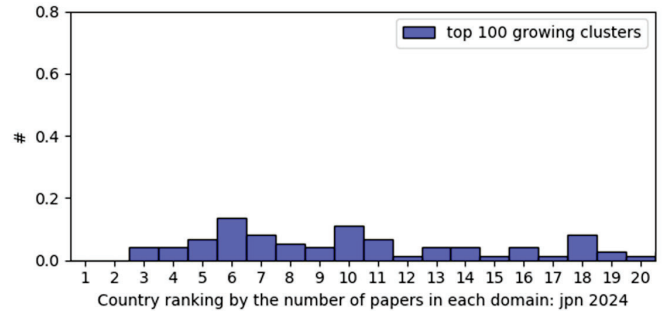
<中国>



<ドイツ>



<日本>

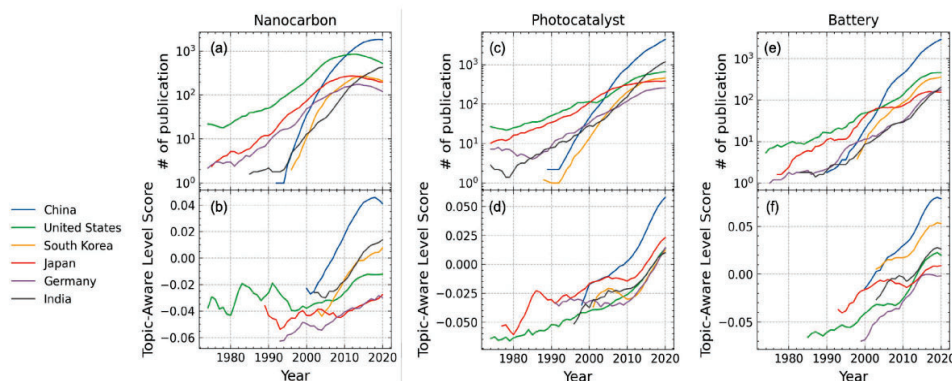


【TI事例②：提案技法・指標の活用】

学術研究の応用移行の早期発見（国別・分野別）

有望な知に関し、基礎から応用さらに市場化へとスピーディに進化させそこで生まれた経済的利益を次世代技術へと積極的に再投資して、世界をリードしようとしている中国。経済成長と同時に研究面でのインパクトも急上昇。

Topic-Aware Level Score (TALS)で測った応用展開度(下図)



※「Topic-Aware Level Score」は、基礎から応用の段階をスコア化したもの。数字が大きい方（図の上）が応用寄り。青線のグラフが中国であり、いずれの分野でも他国に先駆けて応用化が進行している。

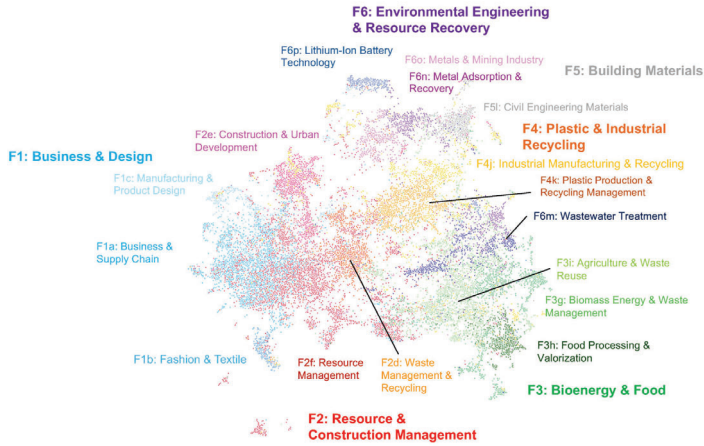
※ISI G20 Innovation & Research Scorecard 2025でも中国のインパクト上昇が強調されている

(Source) N. Higashide, K. Asatani, I. Sakata, "Quantifying advances from basic research to applied research in material science", *Technovation* 135 (2024), 103050.

【TI事例③：提案技法・指標の活用】

融合が必要な新領域における異分野融合の設計支援

サーキュラーエコノミーの学術世界の可視化

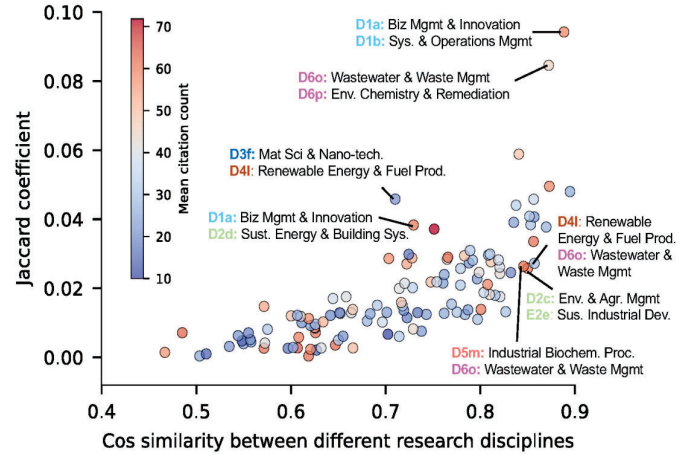


(備考) アブストラクトから生成したセマンティックなVectorsを用いた分類

(Source) K. Nishimoto, K. Kimita, S. Murakami, Y. Long, K. Asatani, I. Sakata, "Advancement of Circular Economy Through Interdisciplinary Collaboration: A Bibliometric Approach", <https://arxiv.org/abs/2507.04923>

必要となる組み合わせの特定

-縦軸は協働が起こる程度-



横軸：領域間の内容的距離、縦軸：分野間の共著割合
色は、当該組み合わせの共著論文群に関する平均の引用数

【TI事例④：技法・指標やノウハウを集約したシステムの提供】

浅谷特任准教授作 Science Planet+ (学術知識の検索システム)

Science Planet+

世界の論文DBのOpenAlex と生成AI等で構築した便利な 学術俯瞰ウェブシステム

<https://planet.sakatalab.t.u-tokyo.ac.jp/>
坂田浅谷研HPより2025.4公開

■約7万の学術領域を網羅

■知識構造マップ

■領域の知識や歴史の解説

■主要な学術的概念

■領域のトップ研究者と 主要な学術論文

■上流・下流の分野との関係

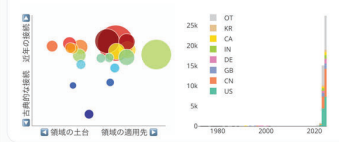
Science Planet+

大規模言語モデルの技術と応用



論文数: 49,872 応用度: 0.578 重要論文 領域内を検索

大規模言語モデル (LLM) は、Transformerアーキテクチャを用いた大規模のテキストから学習し、自然言語生成が可能で、Few-Shot Learningや転移学習により多様なタスクへ適応し、医療や教育などへの応用が期待されています。RLHFやRAG技術で性能を向上させる一方、バイアスや幻覚などの課題の解決にも努めています。LLMの信頼性や倫理的側面を考慮し、持続可能な社会への貢献を目指しています。



トップ研究者
Jason Lee, Morgan Cheatham, Victor Tseng, Tiffany H. Kung, Denny Zhou, Malik Sallam, Maarten Bosma, Xuezhi Wang, Hyung Won Chung, Pamela Mishkin, Aakanksha Chowdhery, Ed H. Joeh, Jochen Kuhn, Stefan Kirchmair, David Charash, Conrad W. Safranek, Jan Leike, Aidan Gilson, Frank Fischer, Kathrin Seifert, Enkeleida Kasneci, Claudia Nierdl, Gjergji Kasneci, Albrecht Schmidt, Michael Sailer, Eyke Hüllermeier, Stephan Güntemann, Stephan Krusche, Georg Groh, Matthias Stadler

トップ研究機関 (日本)
Masashi Reid, Tetsuka Matsuo, Yusuke Iwasawa, Takeshi Kojima, Taro Shimizu, Takanobu Hirose, Yukinori Harada, Ren Kawamura, Zhe Tan, Maiko Takeuchi, Benjamin Heinzerling, Minh Chien Yu, Ken Kawamura, Tetsu Sakamoto, Masashi Yokose, Takahito Watanabe, Soichi Takagi, Kota Sakaguchi, Ayano Ebata, Daiju Ueda, Yukio Miki, Hiroyuki Tatekawa, Shannon L. Walton, Yasuhiro Mitsuyama, Hiroaki Takita, Daisuke Horiochi, Hiroaki Hayashi, Kenkichi Matsushita, Toshimichi Kishimoto, Kazuki Tokumasa

分野の詳細

大規模言語モデル (LLM) は、Transformerアーキテクチャを用いた大規模のテキストから学習し、自然言語生成が可能で、Few-Shot Learningや転移学習により多様なタスクへ適応し、医療や教育などへの応用が期待されています。Chain-of-Thought PromptingやInstructGPTで用いられる人間のフィードバックによる強化学習 (RLHF) といった技術は、LLMの理解力と指示従従能力を向上させました。しかし、LLMはバイアスや幻覚 (hallucination)、知識の欠落といった課題も抱えています。バイアス軽減のためには、学習データの偏りを修正する手法や、FOLIOのようなインスツメンツレベルのプレフィックスを用いたデバッグ手法が研究されています。幻覚対策としては、AlmanacのようなRetrieval-Augmented Generation (RAG) による外部知識活用や、セマンティックエンタロピーに基づく不確実性推定による幻覚検出が提案されています。また、MixtralのようなMixture of Experts 戦略による効率的なファインチューニングも注目を集めました。2023年には、ChainGPTの発表により、大規模言語モデルの社会への影響が顕在化しました。ProGenによるタンパク質生成、HELMによる包括的評価フレームワークの登場なども重要な進展です。医学部附属医科の高いパフォーマンスを示す一方で、偽情報生成やバイアスといった倫理的懸念も指摘され、LLMの出力の正確性と信頼性検証の必要性が高まりました。2024年には、LLMの知識範囲における文脈整合性向上のためのGaussian noiseサンプリングや、幻覚発生メカニズム解明に向けた計算理論に基づく研究が進展しました。医療分野におけるLLMのバイアス定量化や診断性能に関する研究も進展し、大規模言語モデルの社会実装に向けた課題と可能性がより具体的に議論されるようになりました。

分野の歴史

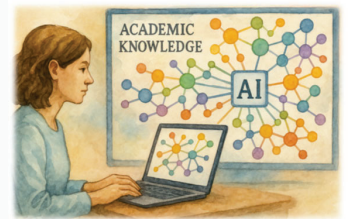
2022年は、Pathwaysシステムを用いたPaLMの登場により、大規模分岐学習によるTransformerアーキテクチャに基づく大規模言語モデルのスケールアップが検証され、Few-Shot Learningの有効性が示されました。Chain-of-Thought PromptingによるLLMの複雑な推論能力が飛躍的に向上し、InstructGPTでは人間のフィードバックによる強化学習により指示従従能力の向上が実現しました。LoRAによるランク分解を用いた効率的なファインチューニングも注目を集めました。2023年には、ChainGPTの発表により、大規模言語モデルの社会への影響が顕在化しました。ProGenによるタンパク質生成、HELMによる包括的評価フレームワークの登場なども重要な進展です。医学部附属医科の高いパフォーマンスを示す一方で、偽情報生成やバイアスといった倫理的懸念も指摘され、LLMの出力の正確性と信頼性検証の必要性が高まりました。2024年には、LLMの知識範囲における文脈整合性向上のためのGaussian noiseサンプリングや、幻覚発生メカニズム解明に向けた計算理論に基づく研究が進展しました。医療分野におけるLLMのバイアス定量化や診断性能に関する研究も進展し、大規模言語モデルの社会実装に向けた課題と可能性がより具体的に議論されるようになりました。

その他の重要概念

GiantRabbit, スパースアテンション, 選択投票メカニズム, XTRUST, モデル監視

頻出語

llms, chatgpt, models, ai, language, data, model, llm, large, tasks, based, human, learning, performance, research, study, knowledge, gpt, using, info, ...



【TI事例⑤:マルチLLMエージェントを用いたアシスタント化】

LLMマルチエージェントによる研究企画支援アシスタント

実務への活用方策として、先のような学術分析手法を実装した専門エージェントに、他の機能を持つ複数のLLMエージェントを組み合わせることで、高度で便利なアシスタントを作成することが可能に。

