

ほらいずん

注目科学技術等の可視化・分析システム — 科学技術の変化の兆しを捉えるツールとして —

科学技術予測・政策基盤調査研究センター 上席研究官 黒木 優太郎

.....

【概要】

本稿ではまず、過去3回分の「専門家が注目する科学技術に関するアンケート調査」のデータ及び同期間3年分の科研費データを比較分析し、文書データの多次元ベクトル化及び次元圧縮によって研究分野マップを作成した結果を紹介する。次に、それらの結果を活用し、注目科学技術と科研費のデータについて外部ユーザが自由に分析できるシステム（注目科学技術等の可視化・分析システム：ST PANGAEA）を整備したので、その主な使い方も紹介する。

今後は、今回作成したシステムを使い、これら未来の科学技術について俯瞰した分析を継続し、さらには各データを継続的にアップデートすることで継続的かつ効率的にホライズン・スキャンニングを行い、科学技術の変化の兆しを捉えていく。

キーワード：科学技術予測, ホライズン・スキャンニング

.....

はじめに

科学技術予測調査について

文部科学省科学技術・学術政策研究所（以下、NISTEP）では1971年から約5年ごとに科学技術予測調査を実施しており、第12回科学技術予測調査を2022年から実施している。2025年5月には、2023～2024年度に行ったデルファイ調査の結果を取りまとめ、調査資料として公表した。第12回科学技術予測調査における最初のステップはホライズン・スキャンニングであり、科学技術や社会のトレンド俯瞰や変化の兆しの早期発見がその目的である（図表1）。

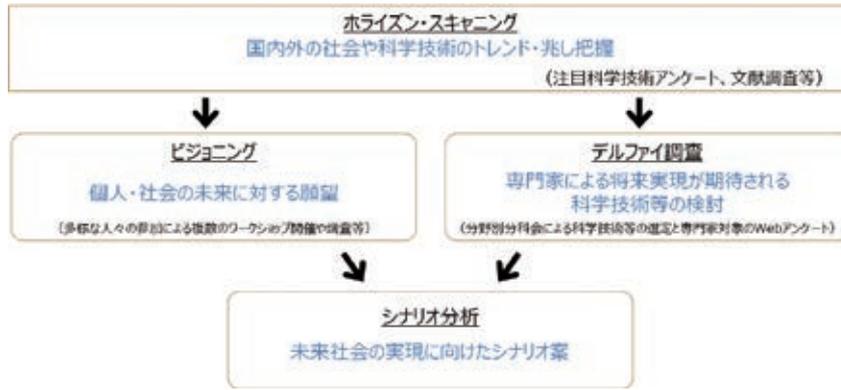
注目科学技術アンケート調査について

注目科学技術アンケート調査は、科学技術予測調査の構成の中で、科学技術のホライズン・スキャンニングに当たる。NISTEPではこれまで、「専門家が注目する科学技術アンケート調査」として、NISTEPが運用する専門家ネットワークに対し、自身が注目する科学技術の概要や実現時期等をアンケート調査し

てきた^{1~3)}。これらの回答は将来実現し得る科学技術等についての専門家による推薦であり、未来に関するデータという点で、学術論文や特許といった既存の情報とは異なる価値を有する。これは、今後30年以内に実現し得る科学技術等を調査するデルファイ調査へつながるものでもある。

これまで約1年おきに調査を実施し、過去3回分の調査データが蓄積されている。第2回調査以降は、自身の分野で今後実現が期待される注目の科学技術を「注目科学技術」、現時点で実現可能性は不明であるものの、将来的に実現すれば学術・経済・社会へ大きな変化をもたらし得る科学技術を「兆し科学技術」として、分けて回答を収集している。これまでの過去3回の調査によって、合計2,351件の回答を得ることができた。これらは未来に関するデータの集合であり、俯瞰・分析することで科学技術の変化の兆しを早期に発見し得る。そのため、注目科学技術の分析補助ツールを設計することは、科学技術予測調査におけるホライズン・スキャンニングの一助となるだけでなく、外部の一般ユーザが科学技術の変化の兆しを探索する場合にも有用であると考えられた。

図表 1 第 12 回科学技術予測調査の構成概要



また、科研費等の研究課題は中長期的研究計画であり、未実現の科学技術という意味では、注目科学技術同様に未来の科学技術について述べられたものであると言える。したがって、注目科学技術だけでなく、科研費も含めた可視化や横断比較が可能なシステムであれば、ホライズン・スキャンニングのツールとして更に強力である。本稿では、実際に設計した注目科学技術等の可視化・分析システム (ST PANGAEA) について、その概要や使い方を解説する。

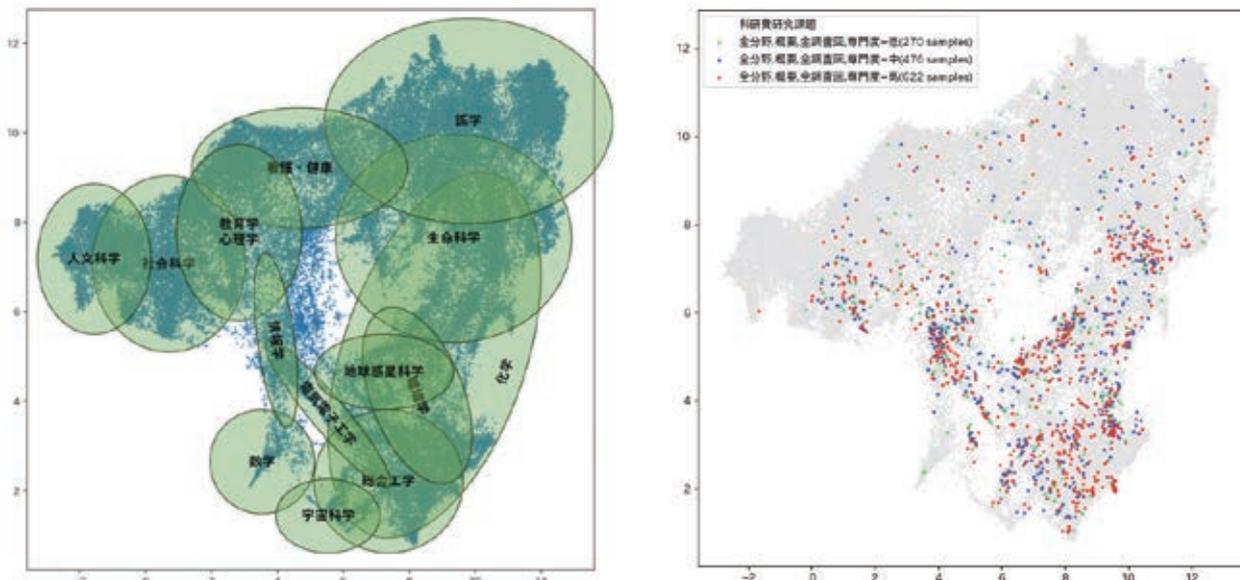
注目科学技術等の可視化・分析システム (ST PANGAEA) の設計

システム作成に当たり、まずは過去 3 回の注目科学技術アンケートの実施時期と同等期間 (2020～2022 年度) の、インターネット上に公開されている科学研究費助成事業のデータ (以下、科研費データ) を可視化して各種マッピングの土台とした。具体的には、対象期間の科研費データ 105,477 件のうち、科

研費の基盤研究 (A・B・C) (応募区分「一般」)、挑戦的研究 (開拓・萌芽)、若手研究で審査区分の小区分又は中区分が設定されているデータ 73,276 件を文書ベクトル化した。ここで「文書ベクトル」とは、MeCab⁴⁾ で文章から各単語を抽出し、分散表現モデルから全単語の多次元ベクトルをそれぞれ抽出し、その平均ベクトルを計算した「文章」単位の多次元ベクトルである。

科研費データの文書ベクトル化においては、研究課題のタイトルと報告書の実施内容などの文章を句点 (。) で連結して一つの文章として取り扱った。ただし、報告書のうち「現在までの達成度」、「今後の研究の推進方策」、「次年度の研究費の使用計画」、「次年度使用額の使用計画」、「備考」の 5 箇所については、研究内容とは関係のない記述が多いため対象外とした。その後、2 次元空間に次元圧縮する UMAP⁵⁾ モデルを作成し、科研費審査の中区分 (65 区分) の研究課題がマップのどこに集中しているかを確認してマップの研究分野を推定した (図表 2、左側)。

図表 2 科研費データマップ (左側) と注目科学技術回答分布 (右側)



続いて、過去の注目科学技術アンケートの全体像を可視化するため、アンケートのうち、注目科学技術 / 兆し科学技術の概要を記述する欄である「概要」の項目を文書ベクトル化し、UMAP モデルを用いて研究分野マップ上に回答者の専門度別にプロットした (図表 2、右側)。

その結果、概要の内容の傾向としては、情報学、生命科学及び理工系全般に対する言及は多数あることがわかった。特に AI に関する内容は回答者の専門分野に関わらず多数見られており、分野に関わらず注目度が高まっていると言える。一方で人文科学や数学に対する言及は特に少なく、教育学、看護・健康及び医学に対する言及も多くはなかった。これは、調査当時の調査対象である専門家ネットワークの専門性の分布によるところも大きい。現在の専門家ネットワークは専門性の偏りに留意しながら運用している。

これまでの分析は、特定の PC 環境や分析スキルを備えていなければ再現が難しい。しかし、冒頭に述べたとおり、科学技術の変化の兆しの探索は一般的なニーズも高いと思われるため、注目科学技術等の可視化・分析システム (ST PANGAEA) を整備し、外部ユーザーが自由に活用できるように一般公開した。本システムは以下 URL からアクセス可能である。

URL : <https://st-pangaea.nistep.go.jp/>

主な使い方

上記 URL にアクセスするとトップページが表示され、「分析例と使い方」から、回答の多かった幾つかの分野を例に、使い方の詳細な解説を見ることが出来る (図表 3)。詳細はそちらを参照いただき、ここで

は主な使い方について概説する。

トップページで「注目科学技術マップから分析する」をクリックすると、実際の UMAP 上に注目科学技術の全回答をプロットした画面が表示される (図表 4)。画面左側に表示されている「データの検索設定」からは、ID 検索や「研究段階」などの調査時の質問項目を対象とした検索が可能であるほか、フリーワードを用いた検索が可能である。フリーワード検索は「部分一致」又は「あいまい検索」が可能であり、あいまい検索の場合には、入力ワードをマップ作成に用いた同技術によって文書ベクトル化して検索する。そのため、自身が入力したフリーワードがマップ上に表示され、マップ上で類似データを探索可能である。「プロット設定」からは、クラスターや調査回など、ユーザが自由に色分け設定及び絞り込みが可能である。

本ツールの主な用途は、単なる検索ではなく、注目科学技術の周辺に存在する類似した学術情報を俯瞰することで、最新の研究動向や科学技術の変化の兆しを把握することにある。そのため本ツールでは、それぞれのプロットをクリックすると、類似したデータ (注目科学技術と科研費の両方) を横断的に検索することが可能である。

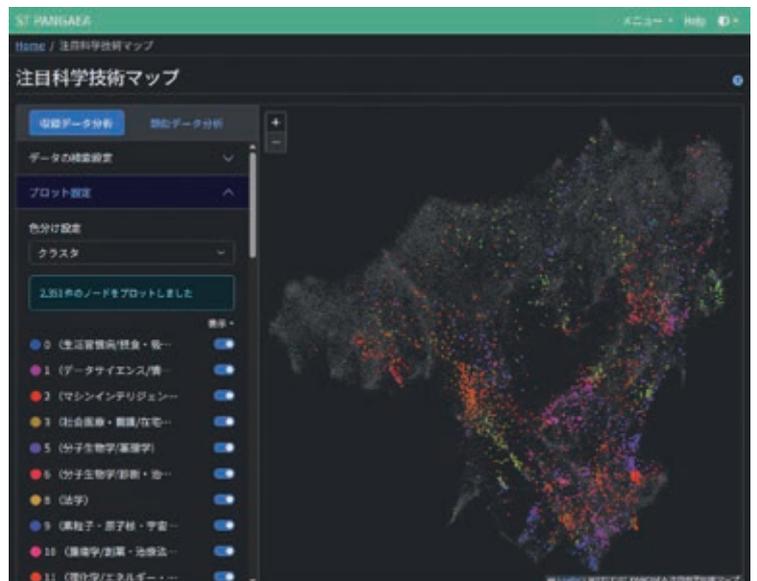
プロットをクリックすると該当データの簡易情報が表示され、ここで虫眼鏡マーク (図表 5、赤枠部分) をクリックすると、類似したデータをプロットさせることができる。これにより、注目科学技術のみならず、その周辺の学術情報を分析することが可能である。また逆に、科研費データから同様の検索をすることも可能である。

また、「クラスタリング結果からマップ分析」や「推定科研費審査区分からマップ分析」では、それぞれの

図表 3 トップページ (イメージ)



図表 4 注目科学技術マップからの分析画面例



図表 5 注目科学技術と類似する研究課題の横断検索



虫眼鏡マーク(赤枠部分)をクリック

図表 6 クラスタリング結果からのマップ分析画面例



クラスタに分類される注目科学技術をプロット

クラスタに分類される研究課題をプロット

分類を起点として注目科学技術又は科研費データを可視化することができる。可視化した後、先述と同様の手順にて、類似データを横断検索することも可能である（図表 6）。

終わりに

本稿では、過去 3 回の注目科学技術アンケートのデータと 3 年分の科研費データを比較分析し、文書ベクトル化や次元圧縮によって可視化した結果を紹介した。また、今後のホライズン・スキャンニングへの活用を念頭に整備した、注目科学技術等の可視化・分

析システム（ST PANGAEA）について、主な使い方を概説した。

注目科学技術はその名の通り、専門調査員が注目する未来の科学技術である。そのため一般的なウェブ検索では情報が不足する場合も多く、今後は本システムを使った効率的調査が可能になったと言える。

また、本システムは拡張性を持たせてあり、注目科学技術や科研費等研究課題に加え、その他のデータを導入することも可能である。今後は、各データをアップデートすることで継続的かつ効率的にホライズン・スキャンニングを行い、科学技術の変化の兆しを捉えていく。

参考文献・資料

- 1) 専門家が注目する科学技術に関するアンケート調査 (NISTEP 注目科学技術 2020), 調査資料 (Research Material), 315 <http://doi.org/10.15108/rm315>
- 2) 専門家が注目する科学技術に関するアンケート調査 (NISTEP 注目科学技術 2022), 調査資料 (Research Material), 325 <http://doi.org/10.15108/rm325>
- 3) 専門家が注目する科学技術に関するアンケート調査 (NISTEP 注目科学技術 2023), 調査資料 (Research Material), 336 <http://doi.org/10.15108/rm336>
- 4) <https://taku910.github.io/mecab/>
- 5) <https://umap-learn.readthedocs.io/en/latest/>