JAIST Repository
https://dspace.jaist.ac.jp/

Title	重点分野分析のための論文マップの活用
Author(s)	寺田, 好秀; 加瀬, 豊; 七丈, 直弘
Citation	年次学術大会講演要旨集, 38: 285-287
Issue Date	2023-10-28
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/19179
Rights	本著作物は研究・イノベーション学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Research Policy and Innovation Management.
Description	一般講演要旨

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

１Ｄ１６

重点分野分析のための論文マップの活用

○寺田好秀（政研大）※，加瀬豊（政研大），七丈直弘（政研大／一橋大）
※yo-terada@grips.ac.jp

1. はじめに

社会課題解決のため、科学技術・イノベーション（以下、STI）政策の重要性が高まっている。また、近年、財政的制約の中で、科学技術関連予算の重点配分が行われている。加えて、政策課題への即時かつ効果的な対応には、エビデンスに基づく政策立案（以下、EBPM）の推進の貢献が期待されている。こうした背景から、内閣府科学技術・イノベーション推進事務局（以下、CSTI 事務局）では、STI 関連データを収集し、データ分析機能を提供するシステム「e-CSTI」の開発を 2018 年度より実施している。この一環として、重点分野分析に資することを目的として、論文書誌情報分析に基づくマップ（以下、論文マップ）とその分析ツールを開発した。本稿では、分析ツールの構想と執筆の時点までに実装された機能を報告する。

2. 先行事例

大規模な論文情報の利用可能性と情報技術の進展に伴い、科学計量学が新しい研究領域としての成立し、論文書誌データベースのインデックスとして付与されている分野情報だけでなく、引用情報に基づく動的な分野分析が用いられるようになってきた。例えば、Henry Small は 1980 年代から引用関係による研究マッピングを開始し、その成果が継承され現在に至るまで、例えば Small, Boyack, and Klavans (2014)のように共引用と直接引用に基づくクラスタリングによる新興分野の同定などが行われている。

このような活動の背景には、近年の科学研究の加速的拡大とグローバル規模で競争が激化した状況がある。ダイナミックに研究分野（特に新興分野や融合領域）の同定や分析に対するニーズが高まった結果、商用サービスとして Elsevier 社の SciVal や、Clarivate 社の InCites などが提供されるに至った。

同様の取り組みは、STI 分野の EBPM の実現のため、いくつかの国内政府系シンクタンクにおいても行われており、手法の開発と分析が実施されている。その先鞭をつけたのは NISTEP の「サイエンスマップ」であり、Web of Science の元データの一つである SCI-EXPANDED を使用し、論文被引用数上位 1%の論文を中心として共引用関係に基づくクラスタリングを行うことで、融合領域や新興領域の把握を可能とした（科学技術政策研究所（2007））。JST-CRDS ではトピックモデルを利用した「俯瞰マップ」により、量子技術の分析を行っている（国立研究開発法人科学技術振興機構研究開発戦略センター（2023））。ライデン大学では、引用関係に基づく論文情報のネットワークを二次元で可視化できる分析ツールとし VOSviewer や CitNetExplorer を作成している。

3. 分析ツールの構築

・論文マップの作成

本稿で報告する分析ツールで活用する論文マップは、依拠する論文書誌情報データベースとして Dimensions を利用している。

Dimensions に含まれる論文には論文単位でその論文の内容を示す Fields of Research（以下、FoR）が付与されている。FoR は ANZSRC による研究活動を分類するためのフレームワークであり、Dimensions では 2008 年版を利用している。Dimensions に含まれる論文には元来 FoR は付与されていないが、Digital Science 社によって機械学習を利用し、論文単位で FoR が付与されている。

被引用数上位 10%論文を対象とした分析ツールの作成においては、各々の FoR (FoR には 1st level と 2nd level があるが、本分析では上位分野を示す 2nd level を利用)において、上位 10%の被引用数を含む論文のみを抽出し、その結果得られた 2010 年から 2019 年までに出版された約 220 万件の論文間の関係を 2 次元空間に投影してマップとした。以上の操作が技術的に可能であることを検証した後に、（被引用数上位 10%に限らない）全論文を対象として同様の分析を行い、全論文マップとした。全論文マップでは 2010 年から 2021 年までに出版された約 2,000 万件の論文が含まれている。以上の処理は、PecanPy(node2vec の実装の一種)等を用いて行われた。

・可視化に用いたソフトウェア

分析ツールには上記のマップのデータを含む書誌情報の可視化とインタラクティブな操作が必要となることから、BI ツールの Tableau を用いて実装した。これは、以下のような CSTI 事務局のニーズに基づく。

まず、CSTI 事務局では分析ツールを CSTI 事務局内で共有して、より多くの行政活動でその利用を促進しようとする計画があった。「分析ツールの共有」を行うためには、共有可能な分析ツールの開発が必要だが、分析内容は多岐にわたり、その内容も年々高度化するため、独自のシステムを 1 から開発するのではなく、BI ツールを活用することで、共有可能な分析ツールの開発にかかる期間とコストの低減が図られた。Tableau を用いることで、e-CSTI のサイトに埋め込まれた Tableau Server を経由した分析ツールの共有が可能となった。また、より高度な分析や分析に要する計算に時間がかかるような処理を行う場合には、デスクトップ版の Tableau を使うなどしている。前者はユーザー毎に異なる権限を与えることで、公開の範囲を細かく制御することが可能である。後者はアクセス制限を行うことはできないが、より複雑な操作を短時間に実施できる。このように、２通りの共有手法を用意し、用途に応じて使い分けている。

・マップ作成の対象となる文献

マップは 2 種作成された。Top10%論文を対象とした分析ツールの論文マップでは、共引用情報を活用して Leiden 法によりクラスタリングを行った。一方、全論文を対象とした分析ツールで活用する論文マップでは、直接引用と文章の類似度により K-means 法を用いてクラスタリングを行った。分析者は、関心のある研究分野に関連するクラスタを特徴語や Tableau を使用した分析ツールを用いて特定することができる。例えば、特定のキーワードが沢山に含まれるクラスタを探すことや、特定の雑誌に掲載された論文が多いクラスタを探ることができる。また、他にも関心のある論文が含まれるクラスタを DOI で検索する機能もある。

e-CSTI の論文マップ、NISTEP のサイエンスマップ、JST-CRDS の俯瞰マップとの比較を表 1 に示す。

表 1　e-CSTI の論文マップ・JST-CRDS の俯瞰マップ・NISTEP のサイエンスマップの比較

	CRDS の俯瞰マップ（量子技術）	NISTEP のサイエンスマップ 2020	e-CSTI の論文マップ	
			Top10%論文を対象とした分析ツール	全論文を対象とした分析ツール
データベース	論文：Scopus 特許：TotalPatent	Web of Science	Dimensions	
論文の出版年	1990 年以降	2015～2021 年	2010～2019 年	2010～2021 年
論文の抽出条件	量子技術に関連するキーワードで検索して抽出した。	各年、各分野において被引用数が上位 1%である Top1%論文を抽出した。	Top10%論文を対象とした。具体的には、出版年ごと FOR ごとに被引用数上位 10%の論文を抽出した。	論文の引用数と被引用数がともに 1 以上で、さらに要約情報を持つ等、論文マップの作成に必要な情報を持つ論文全てを抽出した。
論文などの件数	論文：約 13.8 万件 特許：約 1.4 万件	約 10.1 万件	約 220 万件	約 2,000 万件
マップの作成方法	トピックモデルによる類似度評価により、文章情報を活用して論文と特許の分類を行った。ここで文章情報とは論文がタイトルと要約で、特許が要約である。 その後、多様体学習により二次元マップ化した。	共引用情報を活用して、共引用の度合いの大きさで論文をグループ化した。 二次元マップ化は ScienceMap visualizer を利用した。	共引用情報を活用して K50 に基づき論文ネットワークを構築。Leiden 法によって論文のクラスタリングを実施する。また、Node2Vec を使って引用ネットワークをベクトル分散表現に変換し、t-SNE を使用してそれを 2 次元空間に縮約して分析ツール内でのマップに使用した。	直接引用情報を利用して論文ネットワークを構築する。さらに、各引用・被引用関係について SciNCL を用いた論文アブストラクトのテキストのベクトル表現の類似度を重みとした。K-means 法でクラスタリングを実施。また、Node2Vec と t-SNE を用いて 2 次元空間に縮約し分析ツール内でのマップに使用した。
特徴語の抽出方法	RAKE（Rapid Automatic Keyword Extraction）を利用した。	tf-idf を用いて重要度の高いキーワードを同定した。その後、SciPy の自然言語処理機能を活用し、各単語の語幹を抽出した。	まず、論文の要約情報を対象として Yake を使用し、キーワードを抽出した。次に、tf-idf を用いて重要度の高いキーワードを同定した。その後、SciPy の自然言語処理機能を活用し、各単語の語幹を抽出した。	

（出典）国立研究開発法人科学技術振興機構研究開発（2023）、七丈編・岩崎・寺田ほか（2022, 2023）、七丈・寺田・加瀬（2022）、文部科学省（2023）、政策研究大学院大学（2023）より筆者作成

・計量書誌分析に基づく指標の算出

各分野の特徴を把握するため、論文単位でいくつかの指標を作成した。具体的には「国際性」（当該論文を引用している論文の著者の所属国の集中度）、「融合度」（当該論文が引用している論文の FoR における集中度）、「波及度」（当該論文を引用した論文の FoR における集中度）という三つの指標を、ハーフィンダール・ハーシュマン指数に基づいて定義した。

4. 分析ツールに実装した機能

国際比較

日本の強み等を国際比較により明らかにする機能のニーズがあったため、各クラスタの論文数の大きさを主要国間で比較できる機能を付与した。また、探索的に特定の国の論文数の順位が高いクラスタのみを表示させる機能を設けた。

加えて、関心のあるクラスタに含まれる論文を国際比較したいという要望から、各論文の特徴を主要国ごとに箱ひげ図で可視化する機能を付与した。

総合知の分析の試み

『第 6 期科学技術・イノベーション基本計画』や『統合イノベーション戦略 2023』に重要性が示されている総合知について分析するため、FOR を利用して国別に各クラスタの研究分野の状況を可視化する機能を付与した。

時系列分析

分析ツールには時系列分析の機能を付与した。これにより、特定のクラスタに含まれる論文数の年次変動などを折れ線グラフにより確認できるようになった。

キーワードの一覧表示

クラスタ内の論文のキーワードを一覧的に確認したというニーズがあったため、特定のクラスタのキーワードをワードクラウドで可視化する機能を付与した。

研究資金配分の状況の把握

論文マップに予算執行データを結合し、各クラスタの研究資金配分の概算的な総額を円の規模として示す機能を付与した。また、研究資金配分が大きいクラスタを探索的に特定することを可能とした。さらに探索的に日本の強みの把握や効果的な科学技術関係予算の配分を検討するため、分析ツールにビジネスフレームワークの一つであるプロダクト・ポートフォリオ・マネジメント分析（以下、PPM 分析）の理論を応用する機能を付与した。この機能では各クラスタの位置を示す縦軸を日本の相対的シェア、横軸を論文数増加率とし、円の規模を研究資金配分の総額とするなどし、分野間での相違を比較可能とした。

5. 結語

本稿では、日本の STI 行政の EBPM として効果的かつ効率的な研究開発の推進や、これまでの研究資金配分状況、国内外の研究動向、および日本の強みと弱みについて客観的なデータを用いて分析するため、e-CSTI の論文マップを活用した重点分野分析ツールについてその概要を解説した。特に、『第 6 期科学技術・イノベーション基本計画』および『統合イノベーション戦略 2023』に示された重点分野の分析ツールの CSTI 事務局のニーズに応じて付与した機能と、その利用状況について報告を行った。

Dimensions を利用した分析ツールは、国際比較、キーワードの一覧表示、総合知、時系列分析、研究資金配分状況等についての機能

が付与された。その結果、情報セキュリティ、リチウムイオン電池、太陽電池、サーキュラーエコノミー、歯科インプラントなどの分野の分析に分析ツールが利用され、内閣府の各種公表資料において使用されている。

謝辞

本稿は、政策研究大学院大学が内閣府から委託を受けた令和 4 度度科学技術基礎調査等委託事業「エビデンスシステムの利活用高度化に関する調査」と令和 5 年度科学技術基礎調査等委託事業「エビデンスシステムの利活用高度化に関する調査」の成果の一部である。

参考文献

Aditya, G. and Jure, L. (2016). Node2vec: Scalable Feature Learning FoR Networks. International Conference on Knowledge Discovery and Data Mining, Vol. 2016, pp. 855-864.

Boyack, K. and Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? Journal of the American Society FoR InFoRmation Science and Technology, Vol. 61(12), pp. 2389-2404.

Boyack, K. and Klavans, R. (2020). A comparison of large-scale science models based on textual, direct citation and hybrid relatedness. Quantitative Science Studies, Vol. 1(4), pp. 1570-1585.

Dresbeck, R. (2015). Scival. Journal of the Medical Library Association, Vol. 103(3).

Ostendorff, M., Rethmeier, N., Augenstein, I., et al. (2022). Neighborhood Contrastive Learning FoR Scientific Document Representations with Citation Embeddings. arXiv.org.

Shahdad, S. Y., Sabahatch, A. and Parveez, R. (2016). Architecture, Issues and Challenges of Wireless Mesh Network. International Conference on Communication and Signal Processing, pp. 0557-0560.

Sharma, R. and Djiaw, V. (2011). Realising the strategic impact of business intelligence tools. VINE, Vol. 41(2), pp. 113-131.

Small, H., Boyack, K. and Klavans, R. (2014). Identifying emerging topics in science and technology. Research Policy, Vol. 43, pp. 1450-1467.

Smita, M., Krishnan, H. R. and Ketan, K. (2021). A Literature Survey and Bibliometric Analysis of Application of Artificial Intelligence Techniques on Wireless Mesh Networks. Library Philosophy and Practice, pp. 1-14.

Singh, V. K., Singh, P. K., Karmakar, M., et al. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. Scientometrics, Vol. 126, pp. 5113-5142.

Van Eck, N. J. & Waltman, L. (2014). CitNetExplorer: A new software tool FoR analyzing and visualizing citation networks. Journal of InFoRmetrics, Vol. 8(4), pp. 802-823.

『第 6 期科学技術・イノベーション基本計画』（令和 3 年 3 月 26 日閣議決定）.

『統合イノベーション戦略 2023』（令和 5 年 6 月 9 日閣議決定）.

池田虎三・上田盟子・マークハンゼン（2022）「研究力分析とデータベース」研究力分析タスクフォース『研究力分析に挑む-研究力分析タスクフォース事例集-』pp.36-63.

面和成（2022）「e-CSTI を活用したブロックチェーン分野の研究動向分析」第 2 回ブロックチェーンセキュリティワークショップ企画セッション. https://www.iwsec.org/bws/2022/pdf/omote_open_bws2022.pdf. [アクセス日: 2023 年 8 月 28 日].

面和成・井上瑶子・寺田好秀ほか（2022）「e-CSTI による研究分野分析 ： 情報セキュリティ分野」『研究・イノベーション学会　第 37 回　年次学術大会　講演要旨集』pp.221-226.

河野英子・井上瑶子・上野剛史ほか（2023）「歯科インプラント研究の動向：内閣府 e-CSTI の開発ツールを用いた Bibliometric 分析」日本補綴歯科学会第 132 回学術大会.

国立研究開発法人科学技術振興機構研究開発戦略センター（2023）『論文・特許マップで見る量子技術の国際動向』.

塩次喜代明・小林敏男・高橋伸夫（2003）『経営管理』有斐閣.

七丈直弘・寺田好秀・加瀬豊（2022）「重点分野分析のための論文マップの作成」『研究・イノベーション学会　第 37 回　年次学術大会　講演要旨集』pp.188-191.

寺田好秀・藤田裕二・七丈直弘（2022）「研究生産性の把握～研究者属性と生産性の対比」『研究・イノベーション学会　第 37 回　年次学術大会　講演要旨集』pp.182-187.

内閣府政策統括官（科学技術・イノベーション担当）付（2019a）『研究力の分析に資するデータ標準化の推進に関するガイドライン』 https://www8.cao.go.jp/cstp/evidence/guideline_honbun.pdf. [アクセス日: 2023 年 8 月 28 日].

内閣府政策統括官（科学技術・イノベーション担当）付（2020）「重要科学技術領域の同定に向けた見える化の試みについて-e-CSTI および Science Map2018 の活用を通じた分析-」 https://www8.cao.go.jp/cstp/gaiyo/yusikisha/20201126/siryo3.pdf [アクセス日: 2023 年 8 月 28 日].

内閣府科学技術・イノベーション担当 エビデンス担当（2019b）「EBPM 等の推進に係る取組状況について」 https://www8.cao.go.jp/cstp/gaiyo/yusikisha/20190314-1/siryo1.pdf. [アクセス日: 2022 年 8 月 20 日].

内閣府科学技術・イノベーション推進事務局（2022a）「e-CSTI 等を用いた RFI の試行分析」 https://www8.cao.go.jp/cstp/gaiyo/sip/220331/siryo4.pdf. [アクセス日: 2023 年 8 月 28 日].

内閣府科学技術・イノベーション推進事務局（2022b）「e-CSTI における情報関連人材に関する分析結果について～大規模調査による社会人の知識ニーズの把握と学生の学びのギャップの見える化の試み～」 https://e-csti.go.jp/wp-content/uploads/2023/01/e-csti-fit-report_202209.pdf. [アクセス日: 2023 年 8 月 28 日].

内閣府科学技術・イノベーション推進事務局（2022c）「e-CSTI を活用した日韓の論文動向の事例分析」 https://www8.cao.go.jp/cstp/gaiyo/yusikisha/20221117/siryo2-3.pdf. [アクセス日: 2023 年 8 月 28 日].

内閣府科学技術・イノベーション推進事務局（2022d）「重要科学技術領域の検討に向けた論文情報等の見える化～e-CSTI における情報セキュリティ分野を例とした試行的な分析～」 https://e-csti.go.jp/wp-content/uploads/2023/01/e-csti-security-report_202209.pdf. [アクセス日: 2023 年 8 月 28 日].

内閣府科学技術・イノベーション推進事務局（2023a）「e-CSTI を活用した予算と論文アウトプットの関係性の分析について」 https://www8.cao.go.jp/cstp/gaiyo/yusikisha/20230323/siryo1.pdf. [アクセス日: 2023 年 8 月 28 日].

内閣府科学技術・イノベーション推進事務局（2023b）「総合知の指標について」 https://www8.cao.go.jp/cstp/gaiyo/yusikisha/20230302/siryo2-1.pdf. [アクセス日: 2023 年 8 月 28 日].

文部科学省科学技術・学術政策研究所（2007）『サイエンスマップ 2004　-論文データベース分析（1999 年から 2004 年）による注目される研究領域の動向調査-』 NISTEP REPORT No.100.

文部科学省科学技術・学術政策研究所（2023）『サイエンスマップ 2020-論文データベース分析（2015-2020）による注目される研究領域の動向調査-』 NISTEP REPORT No.196.