

Title	科学技術知識のマッピング研究の現状と今後の課題
Author(s)	伊神, 正貫; 阪, 彩香; 梶川, 裕矢
Citation	年次学術大会講演要旨集, 25: 825-828
Issue Date	2010-10-09
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/9420
Rights	本著作物は研究・技術計画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 Science Policy and Research Management.
Description	一般講演要旨

科学技術知識のマッピング研究の現状と今後の課題

○伊神正貴，阪 彩香（文科省・科学技術政策研），梶川裕矢（東大）

1. はじめに

科学技術の知の構造や発展を客観的に記述する上で計量書誌学の手法は強力なツールとなる。黎明期における Henry Small らによる共引用関係を用いたマッピング研究の発表[1,2]から数十年がたち、近年の劇的な情報処理技術の進展や、科学論文や特許のデータベース整備は、この分野の研究に革新をもたらした。特に知識のマッピングは新たな研究として注目を浴びており多くの研究が欧米を中心に行なわれている[3]。

科学技術知識のマッピング研究は何処まで進展し、現状では何が把握可能なのか。また、今後、科学技術イノベーション政策への適用を考えたとき何が課題となるのか。本講演では、日本におけるマッピング研究の進展の具体例を紹介し、そこから見えてくる科学技術知識のマッピング研究の現状と今後の課題についてまとめる。

2. 日本におけるマッピング研究の例

ここでは、科学技術政策やイノベーション政策への応用という視点から、日本において継続的に科学技術知識のマッピング研究を行っている科学技術政策研究所(サイエンスマップ)と東京大学(学術俯瞰マップ)における研究の具体的な進展について概観する。

2-1. サイエンスマップ研究の進展[4]

科学技術政策研究所において作成しているサイエンスマップは科学研究の動的变化を定期的に観測することを目的に行なわれている研究であり、マッピングの対象を一般的に認識されている分野概念より一段小さい研究領域としている点が特徴である。サイエンスマップを用いた科学研究の分析は、①共引用を用いた論文のグループ化による研究領域の構築、②研究領域のマッピングによる可視化、③研究領域の内容分析の3つを経て行なわれる。過去に4回の研究が行われている。以下に進展の状況をしめす。

(技術予測の一環として調査開始: 2003～2004 年度)

第1回目のサイエンスマップの作成は科学技術予測の一環として実施された。その目的は、急速に発展しつつある研

究領域を抽出する事であり、マッピングの対象は主に個々の研究領域であった。これによりマッピングが研究領域の構造を可視化するのに有効な手段であることが確認された。

(科学研究俯瞰への挑戦: 2005～2006 年度)

続く第2回目のサイエンスマップ2004では、研究領域間の関係を示した研究領域相関マップを導入した。研究領域相関マップから研究領域はマップ上で更なる高次構造(研究領域群)を形成することが明かになり、ナノサイエンスの研究領域群が化学や物理学に關係する研究領域群の間に形成されている事などが明かになった。

(マッピング手法の確立: 2007 年度)

研究領域相関マップによって研究領域が互いにどのような位置関係にあるのかが俯瞰的に把握可能となる。しかし、個々のマップで得られるのは科学研究のある期間におけるスナップショットであり、研究領域間の関係がどのように変化しているのか、これまでは観測されなかった新たな研究領域が生まれつつあるかなどの、科学研究の時系列変化については追跡する事が出来ない。そこで、第3回目のサイエンスマップ2006では、サイエンスマップ2004と2006を接合し、科学研究の時系列分析を可能とするマッピング手法の開発を行った。また、各国の論文シェアなどの統計情報をマップ上に重ねて表示する可視化方法を開発した。

(科学技術イノベーション政策のツールとしての活用方法探索: 2009 年度)

第1～3回調査を通じて科学研究の全体像を俯瞰するためのマッピング手法の基盤が整備された。科学研究を俯瞰するという試みは注目を浴び、サイエンスマップの結果は新聞や学会誌などで多く紹介されている。また、NISTEPによる分析がOECD STI Scoreboardに採用されるなど、国際的な注目も浴びている。一方で、我が国の科学技術政策立案への活用という点では、これまでに目立った活用は行われていない。

第4回目となるサイエンスマップ2008においては、科学技術イノベーション政策のツールとしてのサイエンスマップの活用方法を探索するために、幾つもの試行的な調査を実施した。以下にその内容をまとめる。

(科学研究のダイナミズムの分析)

サイエンスマップ 2002、2004、2006、2008 の4時点のデータを接続することで、科学研究のダイナミズムの分析を実施した。4 時点のデータを時系列比較することで、研究領域が時間の変化とともに、生成、成長、消滅、移動、分割、融合するようすが、マップ上の研究領域の軌跡として表現されることを確認した。また、異なる分野の知識の組み合わせが必要な学際的・分野融合的な研究領域が、2002～2008 年のうちに科学全般に浸透しつつあるようすが観測された。

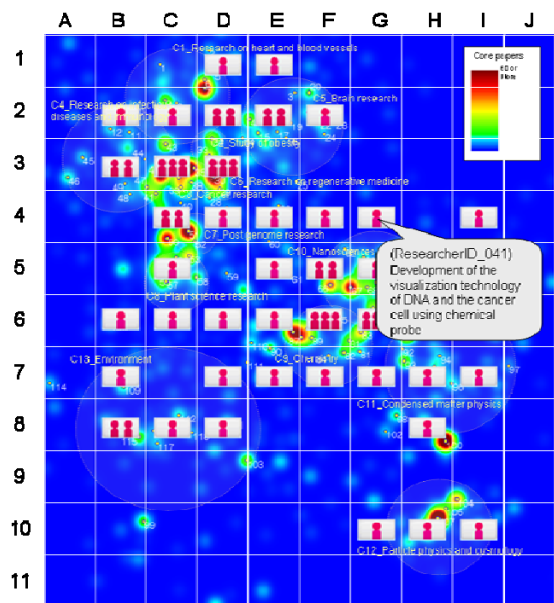
(高い継続性や波及効果を持つ研究領域の探索)

サイエンスマップ 2004 で得られた研究領域のうち、1)どの程度がサイエンスマップ 2008 まで継続しているのか(継続性)、2)幾つがサイエンスマップ 2008 の研究領域に波及しているのか、その広がりはどうか(波及効果)を調べ、高い継続性や波及効果を持つ研究領域を、研究領域を構成するコアペーパー数とネットワーク指標を用いて分類した。

(研究領域内の機関共著ネットワークの可視化)

研究領域内の機関間の共著関係を可視化した。これにより、研究領域の特徴を、機関間の相互関係(研究機関共著ネットワークの密度、国際的な機関共著ネットワークにおける各国機関の位置づけなど)という視点から分析することを試みた。

図表1 インタラクティブサイエンスマップの例: 今後重要となると考えられる研究テーマをサイエンスマップ 2008 上に記入して貰った結果。



データ: Thomson Reuter 社 “Essential Science Indicator”をもとに科学技術政策研究所で分析

(インタラクティブサイエンスマップ)

論文分析による科学の状況把握を行うことは、定量的に分析結果を示すことができる点で優れている。しかしながら、論文の書誌情報や被引用回数などの情報を使うため、近過去

の状況となり、これまでにサイエンスマップ調査分析に参加した研究者からは現時点での状況とのギャップを埋める必要性について指摘を受けてきた。

そこで、従来のように定量的に分析できる部分に加え、「現在研究者が体感し、認識している研究領域の情報」を、研究者に対するウェブ調査を通じて収集することで、近過去から現在・近未来のギャップの補完を試みた。

2-2. 学術俯瞰マップの進展[5]

一方、東京大学の研究グループでは、上記のようなマップを作成するための基盤となるリンクマイニングおよびテキストマイニング技術、ならびに作成されたマップの有効性の評価や、新たな応用可能性の開拓を行っている。

上記のようなサイエンスマップ、学術俯瞰マップの作製は、通常、以下のステップからなる。

(1) 分析対象と目的の設定

まず分析対象となる領域と分析目的を設定する必要がある。領域としては例えば、太陽電池や二次電池のように産業上、国家戦略上の位置づけが大きく、近年、急速に発展している領域や、がん治療のように膨大な量の蓄積が存在する領域、サービスイノベーションのように領域が包括的であるために具体的に何が研究されているのかが見えにくい領域等が挙げられるだろう。

(2) データの収集

対象領域を設定した後に、対象となるデータの種類や収集方法を定め、文書やその書誌情報の収集を行う。論文を対象とするのか、特許を対象とするのか、WWW 上の文書を対象とするのか、また、どのように収集するのか、網羅的にクローリングするのか、検索語を用いて収集するのか、また、データベースは何を用いるのかということを決める必要がある。

(3) ネットワークデータの構築

データを取得後、ネットワークデータを構築する。引用分析では、論文をノード、引用をリンクとみなし、引用ネットワークを生成する。ただし、引用ネットワークの構築方法には複数の方法が存在する。代表的な引用ネットワークの構築方法として、(1) 直接引用 (direct citation)、(2) 共引用 (co-citation)、(3) 書誌結合 (bibliographic coupling)の3つの手法が存在する。

(4) データ分析

代表的な分析手法には、(1)ランキング、(2)クラスタリング、(3)構造予測が存在する。ネットワークとは、ノードとノード間のリンクから構成されるものの総称であるから、原理的には、全てのノードの情報と、全てのノードの組み合わせに対するリンクの情報がそのネットワークの有するデータの全てといえる。

しかし、ネットワークの構造は概して非常に複雑であり、そのような複雑なデータからネットワークの全体像を我々が推測するのは非常に困難である。そのため、ネットワークをクラスタリングなどの方法によって単純化する、もしくはネットワークの特徴量と呼ばれる定量的な指標を用いてノードをランキングするというのが一般的に行われている。

(5) 分析結果の解釈と活用

以上、引用分析の手順と主な分析手法を解説したが、引用分析の最も重要なプロセスは、分析結果をどのように解釈し、それをどのように今後の研究開発マネジメントに、または戦略的な意思決定に活用していくかということである。情報科学の手法が全てそうであるように、リンクマイニングや引用分析は、与えられた情報を分析しているに過ぎない。それを解釈し、分析に価値を与えるのは分析者やユーザ自身である。分析結果の解釈やその有効性・有用性は各自が既に有している知識や経験に大きく依存している。

上記のプロセス全てにおいて、要素技術の開発が進んでいる。例えば、(3)に関して、有望領域をいち早く検出するためのネットワークデータ構築法や、(4)に関して、クラスタリング結果の頑健性の検証、リンクマイニングによる時系列データの予測等が行われている。また、(5)に関しては、マッピング結果の様々な応用可能性が検討されている。例えば、学術マップと特許マップを用いて、先端的な学術研究として実施されているが特許がほとんど取得されていない技術の特定や、NEDOが複数の専門家に依頼して作成した技術ロードマップと比較することで、学術俯瞰マップの優位性を実証した研究が存在する。

4. 今後の課題

これまでに見てきたように、サイエンスマップ、学術俯瞰マップのいずれについても科学技術知識を可視化するための基礎的な手法開発はひと段落し、それらを科学技術イノベーション政策のツールとしてどのように利用できるかを探索しつつある状況であることが分かる。

以降に、科学技術知識のマッピング研究の現状を踏まえた今後の研究課題と、ツールとしての活用における課題についてまとめる。

4-1. 研究課題

(認知空間と社会空間の関係性の分析)

科学技術の状況を一目で俯瞰できるマップは美しく、それ自体で見るものを感嘆させる力を持っている。しかしながら、それは科学技術の知識、いわば認知空間の情報を可視化したものであるため、科学技術イノベーション政策へのインプリケーションには結びつけにくい。

科学技術イノベーション政策に役立つマップとするためには、科学技術の知識とそれが生み出される活動の関係性、つまり認知空間と社会空間の関連性の分析が必要であろう。サイエンスマップ上において、マップ上に各国シェアを重ねて表示することなどが、この例としてあげられる。サイエンスマップ上に示した中国論文シェアの図からは、中国が存在感を見せている研究領域が徐々に増加しているようすが見てとれる。

研究領域における機関や研究者間の協力や競争の状況、研究資金の状況などの情報を加えることが出来れば、研究領域の形成と研究活動の関連が分析可能になるであろう。

研究領域中の機関や研究者ネットワークの分析を正確に行うには、機関や研究者の名寄せが必要となる。一方で、研究資金等の外部情報を加えるには、ミクロレベルでのデータ接続が求められるであろう。

(科学から特許への知識移転測定への応用)

科学と技術の関係という視点で考えるとサイエンスマップで観測できる範囲は、研究の成果が論文と言う形で発表される科学に限られている。科学や技術をより俯瞰的に観測するには、例えば特許などの情報を通じて技術を俯瞰したテクノロジーマップも作成する必要がある。特許文献中で引用されている論文の情報を用いれば、科学と技術のマップをリンクすることも可能となり、科学から技術、技術から科学への知識移転や、その分野依存性も観測できるようになるであろう。

但し、引用情報を用いた特許のマップの作成には困難を伴う。論文とは異なり特許出願については、出願者が権利を得たいと考える各国に出願が行われる。先行技術への引用についても各国の特許出願に対して行われる。また、国によって先行技術への引用数が大きく異なることも知られている。これらの影響を取り除くには、3 極特許ファミリーのように同一の発明についての出願を統合する必要がある。

また、特許における論文への引用情報は、サイエンスリンケージの分析に必須の情報であるが、通常、非特許文献への引用情報はクリーニングされておらず、大規模なデータクリーニングが必要となる。

上記の課題を解決する手法の一つが、東京大学にて開発されているテキストマイニング技術を用いて、論文と特許の距離を測定する手法である。このことにより、引用関係が存在しないもしくは大規模なクリーニングが必要で分析が困難な論文と特許の間の関係性を分析することができる。

(科学技術の状況の実時間観測)

定量的な手法を用いて科学や技術の状況を把握するために、マッピングは強力な手法となりえる。しかしながら、マップから得られる情報は近過去のものとならざるを得ず、最新の科学研究の情報を知りたいという行政ニーズとはかい離があ

る。

このギャップを埋める方法の一つとして、インタラクティブマップのように、ある時点のマップをもとに、科学者の意見を追加していく方法が有効であろう。

更に考えを進めれば、論文データベースとマッピングエンジンを直に連結することで、科学や技術の状況の実時間観測も可能となるかも知れない。

(科学技術予測への応用)

サイエンスマップの時系列研究から明らかになったのは、科学研究の生成、消滅、移動、分割、融合するようすが、マップ上の研究領域の軌跡として表現されることである。知識の進展が、マップ上の力学として表現されるのは驚きである。

研究領域のマップ上での動きを外挿することで、急成長する研究領域や今後関係性を深める研究領域などが予測出来る可能性がある。

今後、その精度を高める必要はあるが、膨大な量の学術情報の中からの急成長している領域の検出、研究領域間の関係性の分析、将来多くの引用数を獲得するであろう有望論文の予測などが既に行われている。

4-2. ツールとしての活用における課題

(科学者や機関が自己のポジションを把握するためのツールとしての応用)

サイエンスマップ、学術俯瞰マップとも、それに対する認知は徐々に広まりつつある。しかしながら、科学者がそれらを積極的に活用するには到っていない。

東京大学総合研究機構イノベーション政策研究センターでは、学術論文の全体像を俯瞰できるマップを自動的に生成できるシステム学術俯瞰マップ作成システムを開発している。科学者がこのようなシステムを活用することで、自らの研究テーマやその周辺の研究の進展の状況を容易に把握できるようになるであろう。

また、サイエンスマップ上に、科学者が入力したキーワードや機関の位置を示すことによって、科学者や機関が自己のポジションを把握するためのツールとしての活用も可能であると考えられる。

これらについては、これまでに得られた可視化手法を活用することで、容易ではないが実現可能であり、科学技術知識のマッピング研究の成果を活用し、その認知度を高めるには良い方法と考えられる。

(乱用や誤用を防ぎ、ツールを使いこなす力を関係者が身に付ける事)

認知度を上げるということは、我々分析者の手から離れて活用されることとなる。データの特性からどこまでが言えて、どこが限界か。他のデータも同時に扱う場合、比較してよい

のか、結合してもよいのか。ツールの有用性を説き認知度を上げるとともに、限界についても我々は提示していく責任が生じる。

乱用や誤用の防止策として、まずは小さなことだが、何のデータベースを用いて、どのような計算方法で分析したかを注釈として明記することを徹底することから始めるのはどうだろうか。活用者も一手間を惜しまず、専門家への確認を行ってほしい。

また、今後より一層多くなる情報(シェアやランキングなど)に踊らされることが無いように、関係者に対してある程度データを読む力を教育していく機会を持つべきであろう。それぞれの立場のコンテキストの中で有用である情報を選択し、神経を使えばよいのである。ユーザがツールに振り回されるのではなく、ツールを活用出来るよう、注意喚起を常時おこなう必要がある。

(機密性と開示性についての問題)

これはマッピング研究に限った事ではないが、科学技術イノベーション政策のツールについては、そこから得られた結果やツールを何処まで公開するかが、将来的には問題になると考えられる。

科学技術の状況の実時間観測や予測の結果は、政策を決定する際の要となる情報であり、機密性が重要となる。一方で、科学としての進展を考えるのであれば、結果やツールは広く使用されることが望ましいであろう。

機密性と開示性、この両者のバランスについても、科学技術イノベーション政策のための科学を推進する上で考えていく必要があるであろう。

参考文献

- [1] Small, H. and Sweeney, E. (1985a), Clustering the Science Citation Index using Co-citations. I. A Comparison of Methods, *Scientometrics*, 7 : 391-409.
- [2] Small, H., Sweeney, E., and Greenlee, E. (1985b), Clustering the Science Citation Index using Co-citations. II. Mapping Science, *Scientometrics*, 8 : 321-340.
- [3] Börner, K., Chen, C., and Boyack, K. W. (2003), Visualizing Knowledge Domains, *Annual Review of Information Science and Technology*, 37 : 179-255.
- [4] 科学技術政策研究所による一連の研究については以下を参照のこと(<http://www.nistep.go.jp/index-j.html>).
- [5] 東京大学大学院工学系研究科総合研究機構イノベーション政策研究センターによる一連の研究については以下を参照のこと(<http://www.ipr-ctr.t.u-tokyo.ac.jp/>).