

論文データベースを用いたサイエンスマップ作成と 研究領域の動向分析

○阪 彩香（文科省・科学技術政策研），伊神正貫（OECD），
桑原輝隆（文科省・科学技術政策研）

1. 研究の目的

現在施行中の第3期科学技術基本計画では、科学技術の戦略的重点化の中の基礎研究の推進という項目がある。文書では、基礎研究には二種類あり、多様性の苗床の形成を促す研究者の自由な発想に基づく研究と、非連続的なイノベーションの源泉となる知識の創出としての政策に基づき将来の応用を目指す基礎研究が掲げられている。科学技術が社会へのインパクトを実現するには、その基盤としてこれらの基礎研究が重要な役割を果たす。エビデンスベースの議論のもと、今後の科学政策が行なうためには、基礎研究の現状を俯瞰的に把握できる資料が必要となる。

そこで我々は、基礎研究を中心とする科学の動向を俯瞰的に把握することの出来る「サイエンスマップ」の作成を試みた。また、今後の科学推進への問題探索やその解決糸口を見出すことへ「サイエンスマップ」を活用することができるかを検討した。

2. 研究の概要

2-1. 手法について

「論文データベース分析を用いた研究領域の構築（以下、論文データベース分析）」と「構築された研究領域に対する内容分析（以下、内容分析）」を組み合わせることで、研究領域の客観的な把握を試みた。論文データベース分析の出発点として、1999年から2004年までの6年間に発行された論文（以後2004年データと記述）の中で、各年、各分野（臨床医学、植物・動物学、化学、物理学など22分野）の被引用数が上位1%である高被引用論文（約4万7千件）を用いた。これら高被引用論文に対して、「共引用」を用いた論文のグループ化を2段階で行うことで、一定の大きさを持つ133の研究領域を構築した。また、得られた133研究領域について、内容分析を行なった。本稿で用いた論文データベースの分析から研究領域の構築を試みる手法は、独自に開発した手法（参考資料1）を改良したものである。

本調査では第1段階のグループ化で得られる論文の集合として Thomson Scientific 社の Essential Science Indicators (ESI) に収録されているリサーチフロント（5350 リサーチフロント）を用いた。ここで「共引用」とは、2つ以上の論文が1つの論文に同時に引用されることを指す。頻繁に共引用される論文は、その内容に一定の共通点があると考えられる。従って、共引用によって結びつけられる論文をグループ化する事で、研究内容に共通性のある論文群を得ることが出来る。本稿では、これらの論文群を「コアペーパー」（研究領域の核を構成する論文）、コアペーパーを引用する論文を「サイティングペーパー」と呼ぶ。

2-2. 133研究領域について

○ 論文データベース分析から得られた133の研究領域について

論文データベース分析により、133の研究領域が構築された。133研究領域を構成するコアペーパーの分野分布を比較し、似た分野分布を持つ領域間に引力が働くモデル（重力モデル）を用いて、サイエンスマップを作成した（図1）。サイエンスマップをみると、42領域が臨床医学や植物・動物学といったライフサイエンスの領域であった。この内、約半数の27領域が臨床医学の領域である。また、物理、化学、工学、材料科学の領域（物理・化学系領域）が45領域、環境/生態学、地球科学の領域が7領域抽出された。少数であるが、宇宙科学、社会科学・一般の領域も抽出された。また、133研究領域の約3割である39領域が学際的・分野融合的領域であることが明らかになった。

○ 研究領域に見る日本の研究活動の特徴

133研究領域の各々について、領域を構成するコアペーパー中の日本論文シェアを見ると、物理学、化学、植物・動物学、材料科学の研究領域では、9%¹以上のシェアを示す領域が多数あり、日本の存在感が大きいことが分かった。一方、工学、環境/生態学、宇宙科学の研究領域においては、9%以上のシェアを示す領域が少なく、日本の存在感は小さい。シェアが0%の研究領域が2割強あり、現在行なわれている研究のバラエティがやや劣ることが分かった。

¹ 133領域を構成する全てのコアペーパーに占める日本論文の割合は約9%。

○ 研究領域の時系列変化

1997 年から 2002 年までの6年間に発行された論文を対象に行なった調査(参考資料1)のデータを用いて、研究領域の時系列分析を試みた。その結果、個々の研究領域の変遷として、①拡大が見られる研究領域、②継続的な進展が見られる研究領域、③安定的な研究領域の3パターンに分類できた。また、上記のように、個々の研究領域の変遷の上に、研究領域全体を捉えてみると、新たに研究領域が生まれる(④新たに見出された研究領域)や研究領域が融合する(⑤融合が見られる研究領域)という事象があると考えられる。2003年に世界を震撼させた重症急性呼吸器症候群(SARS)やトリインフルエンザの発生に伴い、これらに関する研究が急速に進展されたことがうかがえる。このように、本手法を用いることで過去1~2年間に形成された注目度の高い研究領域の把握も可能である。さらに、融合が見られる研究領域をみると、②継続的な進展が見られる研究領域が多いことが分かった。

そして、研究領域のマップを比較することで、時系列でどのような変化が起きたのかを可視化することができることが分かった。例えば、バイオテロに関する研究領域においては、菌そのものの研究から、実際の緊急時にどのように対応すればよいかという社会的視点を踏まえ進展していることが明らかとなった。

加えて、サイティングペーパーの時系列変化の分析から、発展領域の内容が時間と共に動的に変化する様子を記述することができた。例えば「マイクロアレイデータの処理に関する統計手法の開発について」の研究領域では、コアペーパーでは数学が主分野であるが、サイティングペーパーをみると19の分野に影響していくことが分かった。

○ 現在の科学活動について

得られた133の研究領域をもとに、研究領域間の共引用関係を用いて、現在の科学活動についての俯瞰的な理解ができるように、研究領域間の相対的な関係を示したマップを作成した(図2)。マッピングの結果と内容分析の結果、133の研究領域は14の研究カテゴリーに分類できることが明らかになった。

マップの上中央に見られる研究領域は、ライフサイエンスもしくは保健・医療に関連した研究領域である。研究領域が複雑に関係しているために、カテゴリーを明確に分類する事は容易ではないが、マップの上半分が保健・医療、マップの下半分がライフサイエンスに関連した研究領域と考えることが出来る。ポストゲノム研究と感染症・免疫研究の研究領域はマップ中に広く広がっている。

マップの中央右にはナノテクノロジー・材料に関連した研究領域が存在する。マップ上の研究領域の位置は、共引用の強さのみに依存しているため、化学合成と超伝導・量子コンピューティングの間に位置する中間的なナノ材料の位置は、ナノ材料研究の境界的・分野融合的な特徴を示している。

素粒子・宇宙論に関連した研究領域がマップの右下に広がっている。他のカテゴリーから離れて位置することから、素粒子・宇宙論の研究は他のカテゴリーとの相互作用が小さいと考えられる。別の言い方をすると、素粒子・宇宙論から他の研究への直接的な知識の流れは非常に限られており、このカテゴリーの研究は伝統的な分野に軸足を持つ研究者の参加によって形作られていると予想される。

環境のカテゴリーはマップの中央部に位置している。これは本カテゴリーがライフサイエンス、保健・環境とナノテクノロジー・材料の両方と一定の共引用関係を持つためと考えられる。ここには、環境変動、化学物質、生物多様性などの研究が含まれる。本カテゴリーには全部で8個の研究領域が含まれている。

本調査は、主に自然科学を対象としているが、分析の結果、ベンチャーキャピタル、新時代の知的財産権といった社会科学に関連する研究領域も得られた。これらの研究領域はマップ上の中央左に位置している。「認知心理現象の脳機能イメージング」、「視覚的注意と脳活動の振動現象」と言った研究領域はライフサイエンスと社会科学の間に位置する境界的・分野融合的領域であることが、マップ上のこれらの領域の位置から分かる。

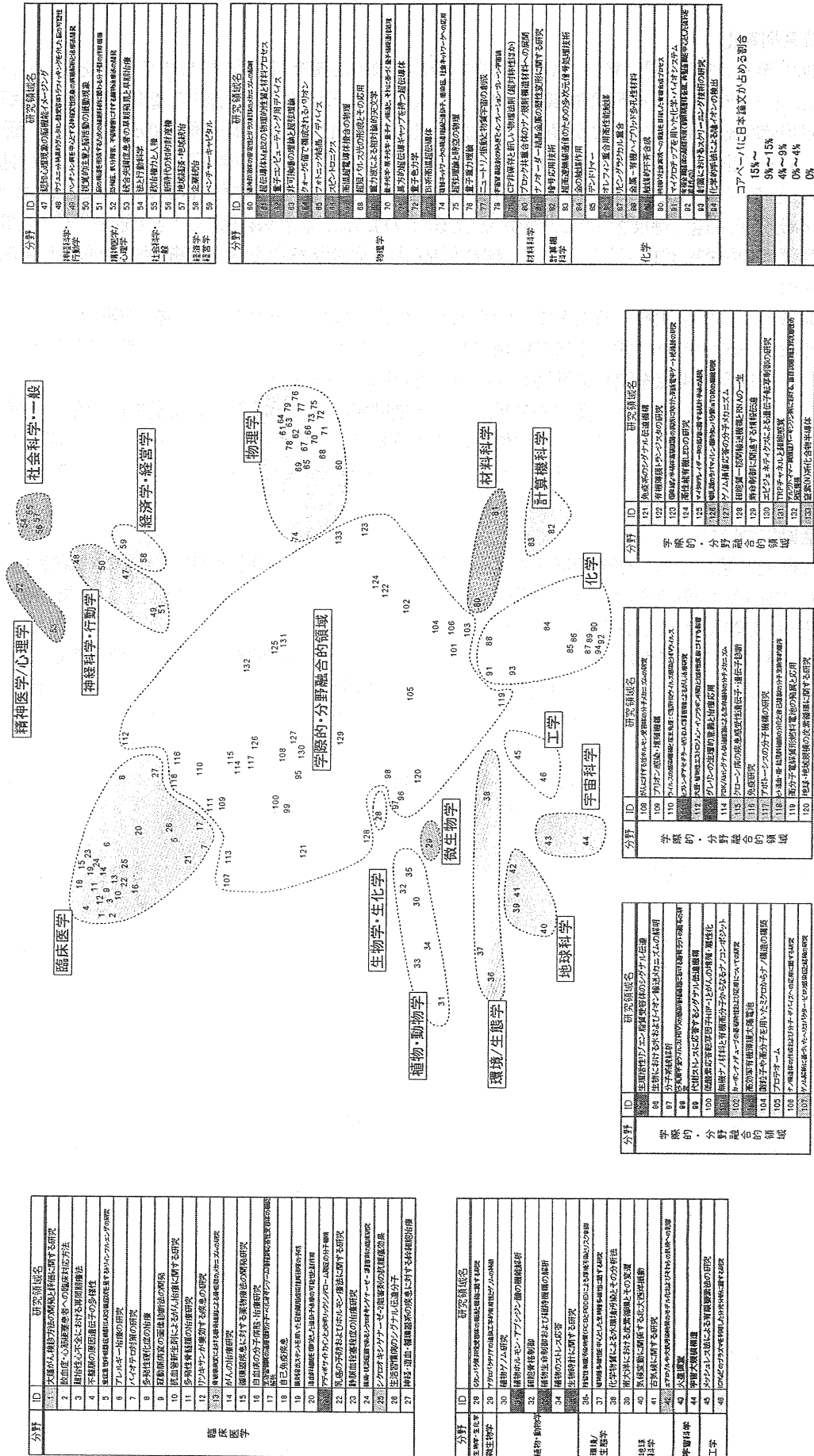
3. まとめ

我々は、基礎研究を中心とする科学の動向を、論文データベース分析によって研究領域レベルで把握し、サイエンスマップを作成する手法を構築した。加えて、研究領域における日本の存在感の分析を通じて、研究領域における日本の研究活動の状況の把握や、個々の研究領域の変遷を可視化することが出来た。

[参考文献]

1. 科学技術政策研究所、「NISTEP REPORT No.95 急速に発展しつつある研究領域調査」2005年5月

図表 1 既存の学問分野と133の研究領域の関係マップ



○ 研究領域を構成するコアペーパーの 22 分野の分布を比較し、似た分野分布を持つ領域間に引力が働くモデル(重力モデル)を用いて作成した。

○ マップ内の番号は、研究領域番号である。

図表 2 研究領域133のカテゴリーマップ

