

Title	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
Author(s)	伊神, 正貫; 阪, 彩香; 桑原, 輝隆
Citation	年次学術大会講演要旨集, 23: 578-581
Issue Date	2008-10-12
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/7629
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

科学研究の時系列分析を可能とするマッピング手法の開発

○伊神 正貴、阪 彩香、桑原 輝隆(文科省・科学技術政策研)

1. はじめに

科学技術の知の構造や発展を計量書誌学の立場から記述する試みは古くからなされている。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[1]。マッピングの対象は多様であり、ジャーナルの引用関係を用いた分野間の関係についての分析、国や組織間の共著関係の分析、研究者間の共著の分析などが研究されている。

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注目研究領域の内容分析の3つを経て行なわれる。

これまでに2回の先行研究が行なわれた。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[2]、[3]、[4]。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。本先行研究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。しかし、この研究では研究領域間がどのような関係にあるのかまでの分析には到らなかった。

続く研究では、研究領域間の関係を示した研究領域相関マップ¹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[5]、[6]。

研究領域のマッピング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、これまでに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、本研究では異なる期間のマップを接合し、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。ここでは主に手法について議論し、具体的な分析例については発表2B08, 2B09で示す。

2. 分析手法

(分析に用いたデータ)

論文のグルーピングに用いた基本データセットは高被引用度論文である。これらは各年、各分野(臨床医学、植物・動物学、化学、物理学など22分野)において被引用数が上位1%に入る論文である。本研究では1999年～2004年、2001年～2006年の2期間を分析の対象とした。前者がサイエンスマップ2004の基本データセットであり、後者がサイエンスマップ2006の基本データセットである。

高被引用度論文間のリンケージについては、 $N_{\text{norm}}^{AB} = n_{AB} / \sqrt{n_A n_B}$ から計算される規格化された共引用度を用いた。ここで、 $n_{A(B)}$ は論文 A(B) の被引用数であり、 n_{AB} は論文 A と B が共引用される回数である。2段階のグルーピングの内、1段階目のグルーピングで得られた高被引用度論文のグループを Research Front と呼び、Research Front を構成する高被引用度論文をコアペーパーと呼ぶ。1段階目の結果については、トムソン・ロイター社の Essential Science Indicators に含まれる Research Front を用いた。データベースには 5,538(5,350)² の Research Front が含まれている。2段階目のグルーピングでは、Research Front³ を1つの仮想的な論文と考え、Research Front のグルーピングを行う。2段階目のグルーピングによって 687(626)の研究領域が得られた。

マッピングは2段階目のグルーピングで得られた全ての研究領域を用いて行った。687 研究領域のうち 124 注目研

¹ 以降では特に断りが無い限り研究領域相関マップをサイエンスマップと呼ぶ。

² カッコ内はサイエンスマップ2004の作成に用いた基本データセットの値を示す。

³ Research Front は2件以上の高被引用度論文から構成されている。

究領域については、その内容分析を専門家に依頼した。各研究領域の詳細については参考文献[7]に記述されている。

(並列マッピングによる科学研究の時系列分析)

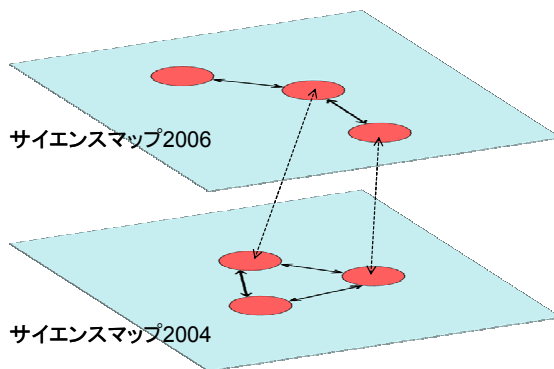
本研究では異なる期間のマップを接合し、科学研究の時系列分析を可能とするマッピング手法の開発を行った。以下にその詳細を示す。ここでは2期間のマップの接合例について述べるが、ここで示す手法は3期間以上へ容易に拡張できる。

1期間のみを対象とした通常のマッピングでは、共引用関係で結び付けられる研究領域間(A-B)に、 $N_{\text{norm}}^{\text{AB}} \times r_{\text{AB}}$ の引力が働くとする。ここで、 $N_{\text{norm}}^{\text{AB}}$ は規格化された共引用度であり、 r_{AB} はマップ上における研究領域間の距離である。引力とは別に斥力も導入される。これはマップがつぶれる事を防ぐ為に導入される力であり、全ての研究領域間に斥力が働くとした。斥力の値は $N_{\text{norm}}^{\text{max}} / r_{\text{AB}}^2$ で計算される。 $N_{\text{norm}}^{\text{max}}$ は規格化された共引用度の最大値、 r_{AB} はマップ上における研究領域間の距離である。この引力、斥力のバランスによりサイエンスマップ上の研究領域の位置は決定される。

上記に述べた引力、斥力のみを用いて、異なる2期間のサイエンスマップを作成することが可能である。しかし、この場合2つのマップは独立であり、互いの対応関係は不明である。そこで本研究では、異なる2期間のマップ間(今回の場合、サイエンスマップ 2004 とサイエンスマップ 2006 の間)に仮想的な引力を導入することで、2つのマップを1つのマップとして取り扱うようにした。具体的には、それぞれのマップを構成する研究領域(例えば D と E)で、共通のコアペーパーを持つものの間に $C^{\text{DE}} \times r_{\text{DE}}$ の引力が働くとした。ここで、 $C^{\text{DE}} = m_{\text{DE}} / \sqrt{m_{\text{D}} m_{\text{E}}}$ である。 $m_{\text{D(E)}}$ は研究領域 D(E)を構成するコアペーパー数、 m_{DE} は研究領域 D と E で共通のコアペーパー数である。 r_{DE} はマップ上における研究領域間の距離である。

2つの引力を模式的に図2に示す。サイエンスマップ 2006 の研究領域は共引用で関係付けられた研究領域との間に引力を感じる(図2中の実線矢印)他に、サイエンスマップ 2004 の研究領域の中で共通のコアペーパーを持つ研究領域との間にも引力を感じる。これによって過去の研究領域からの履歴と現在の研究領域間の関係を同時に考慮した形でマッピングが可能となる。

図2 並列マッピングの模式図



(マップの可視化方法)

通常、マッピングからは研究領域の中心位置のみが得られる。従って、マップ上で研究領域の広がりやをどのような方法で表現するかについては自由度が残る。

今回、サイエンスマップでは研究領域の広がりをガウス関数で表現する事とした。これは2つの要因による。1つめの要因はマップと地形地図とのアナロジーを出すというデザイン上の決定である。2つめの要因は円による表現の技術的な問題である。研究領域に含まれるコアペーパー数が最大/最小の比が 10^4 のオーダーとなる為、円の大小でマップを表現する事は困難となる。

具体的には、研究領域のコアペーパーがマッピングプログラムで求められた位置を中心にガウス関数状に分布しているとした。サイエンスマップにおけるコアペーパー分布は、各研究領域に対して得られるガウス関数の重ね合わせ $\rho(x, y)$ (サイエンスマップ 2006 においては 687 研究領域) で表現される。 $\rho(x, y)$ を全平面について積分した値が全コアペーパー数となる。

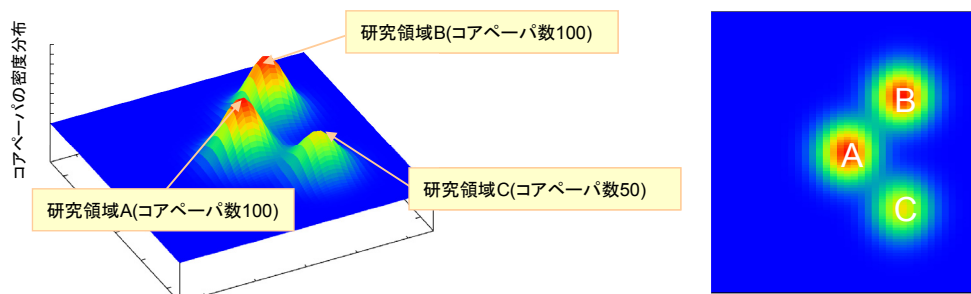
実際の可視化の際には、コアペーパーの分布を、コンピュータを用いて2次元平面に表現するために、サイエンスマップを面積 $dx \times dy$ のメッシュに分割した。次に、メッシュ内に含まれるコアペーパー数をカウントし、コアペーパーの密度

(コアペーパー数/($dx \times dy$))で $\rho(x, y)$ を近似した。研究領域相関マップでは各メッシュをコアペーパー密度に対応する色で色づけしている。

図 2 マップの可視化方法の模式図

(a)コアペーパーがガウス関数状に分布する様子

(b)真上から図 2(a)を見たもの



3. サイエンスマップ 2004 と 2006

並列マッピングによって得られたサイエンスマップ 2004 と 2006 を図 3 に示す。サイエンスマップからは研究領域が幾つかの研究領域群に分かれ、これらの研究領域群が互いに関係していることが分かる。細かな位置関係は異なるが、研究領域群の大まかな位置はサイエンスマップ 2004 と 2006 で概ね同じであり、新たに導入したマップ間の引力が作用している事が分かる。

サイエンスマップ 2004 と 2006 の大きな構造に注目すると、サイエンスマップの右下の研究領域は、素粒子・宇宙論に関するものである。その上には、物性研究の研究領域群が広がっている。相関マップの中央には、大きく分けて2つの研究領域群が存在する。中央右側はナノサイエンスの研究領域群であり、中央左側には化学合成に関連した研究領域群が広がっている。化学合成の左下には、環境に関連した研究領域群が存在する。これらは、化学合成の研究領域のように一箇所に集中するのではなく、マップ上で広がりを持って存在している。

サイエンスマップの左上部分には生命科学にかかわる研究領域群が集まっている。一番下の化学合成に近い部分に存在するのが植物科学研究である。その上にポストゲノム研究が広がり、感染症・免疫研究、がん研究、肥満研究、脳研究へと繋がっていく。一番上の部分に広がっているのは、心臓・血管疾患にかかわる研究領域群である。

4. まとめ

本研究では、科学研究の時系列分析を可能とするマッピング手法を開発した。同じ期間のマップ内の研究領域間のリンケージに加えて、異なる期間の研究領域間のリンケージを考えることで、過去の研究領域からの履歴と現在の研究領域間の関係を同時に考慮した形でマッピングが可能となった。

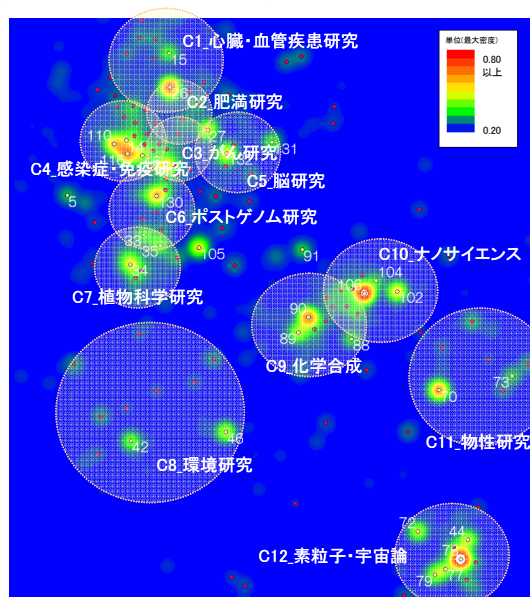
この方法の3期間以上への拡張は容易である。例えばサイエンスマップ 2008 を新たに接続したい場合は、サイエンスマップ 2006 と 2008 の間に引力を導入しマッピングを行えば良い。

(参考文献)

- [1] Börner, K., Chen, C., and Boyack, K. W. (2003), Visualizing Knowledge Domains, Annual Review of Information Science and Technology, 37 : 179-255.
- [2] 伊神正貴, 桑原輝隆, 論文データベースを用いた新興科学技術領域の俯瞰的探索手法, 研究・技術計画学会第 18 回年次学術大会, 2003 年 11 月
- [3] 伊神正貴, 阪 彩香, 桑原輝隆, 論文データベースによる研究領域の俯瞰的探索, 研究・技術計画学会第 19 回年次学術大会, 2004 年 10 月
- [4] 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
- [5] 阪 彩香, 伊神正貴, 桑原輝隆. 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, 研究・技術計画学会第 21 回年次学術大会, 2006 年 10 月
- [6] 科学技術政策研究所, NISTEP REPORT No.100 サイエンスマップ 2004, 2007 年 3 月
- [7] 阪 彩香, 伊神正貴, 桑原輝隆, NISTEP REPORT No.110 サイエンスマップ 2006, 2008 年 6 月

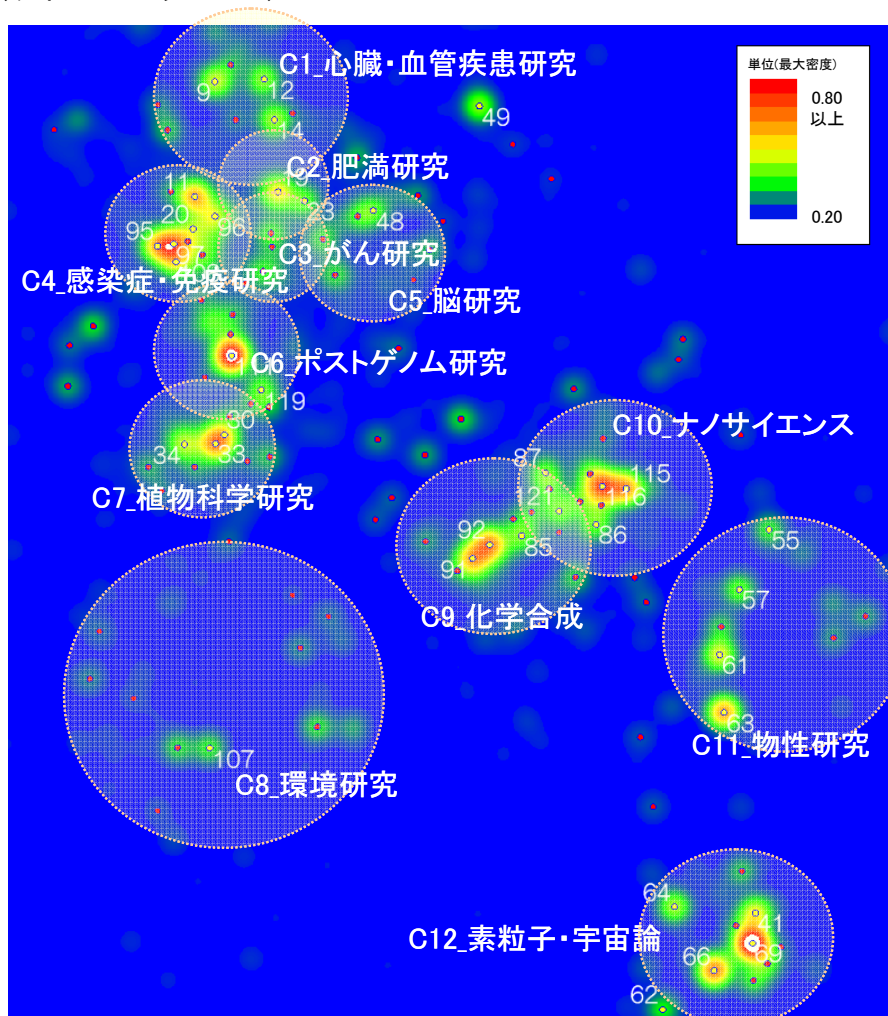
図 3 サイエンスマップ 2004 と 2006

(サイエンスマップ 2004)



- 黄色の丸が注目研究領域の中心位置を示し、丸の横に書かれた数字は注目研究領域の ID を示す。マップ中のグラデーションはコアペーパーの密度に対応している。コアペーパーが集中している部分は暖色、コアペーパーの密度が小さくなるにつれ色が次第に寒色に近づく。色はサイエンスマップ中で最もコアペーパーの密度が高い部分を基準として決められている。最大密度の 0.8 倍の密度を持つ部分は赤色、それ以上の部分は白抜きで表現している。
- 他研究領域との共引用が小さい一部の領域は、マップの中心から外れた位置に存在するため、上記マップには描かれていない。

(サイエンスマップ 2006)



データ：トムソン・ロイター社“Essential Science Indicators”に基づき筆者が集計。