

Title	共語分析の手法を用いた学術研究の科学技術への波及に関する評価手法の開発
Author(s)	橋本, 泰一; 内海, 和夫; 近藤, 隆; 石川, 正道
Citation	年次学術大会講演要旨集, 24: 267-270
Issue Date	2009-10-24
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/8625
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

共語分析の手法を用いた 学術研究の科学技術への波及に関する評価手法の開発

橋本泰一、内海和夫（東京工業大学）
近藤 隆（三菱総合研究所）、○石川正道（東京工業大学）

1. はじめに

D.E.ストークスが提案した「パスツール象限」の考え方に代表されるように、自然の根源的理解と実用化への意識が共存する科学技術の領域がある¹。半導体、バイオ、医療など、サイエンス型産業とも呼ばれる領域である。このような分野では、自然現象の探究と実用性の追求が極めて近い関係にあり、科学研究の成果が直接新技術へと波及する。パスツール象限は、米国において大学の研究が新しい製品や製造方法を生み出したことに関する実証的な研究を背景として、基礎研究に公的資金を拠出する理由を明快に説明する考えとして広く知られるようになった。しかしながら、我が国における実態の把握となると、基礎研究と技術進歩とを計量的に結びつける研究は、玉田ら（2002）による先駆的な研究があるものの、乏しいと言わざるを得ない²。

玉田らは、特許1件あたりの引用論文等の数、すなわちサイエンスリンケージを、科学技術基本計画において重点分野とされるバイオテクノロジー、ナノテクノロジー、IT及び環境の4分野について、日本の特許を対象とした計測を実施した。その結果、日本特許のサイエンスリンケージは、技術分野によって大きく異なり、バイオテクノロジーが突出して多く、ナノテクがこれに続き、ITと環境技術は少ないという一定の傾向をもつことを明らかにした³。サイエンスリンケージは、特許申請書に明確に引用される非特許引用文献（論文等とよぶ）の数を、科学の指標として計量する方法である。この定義によると、工学など学問分野化した技術も、その成果物が論文という形をとる限りは科学に含まれることに注意が必要である。サイエンスリンケージの方法を日本特許に適用する場合の課題は、特許法において米国特許と異なり、申請された技術の新規性を立証するために、関連する特許や非特許文献を出願人が記載する法的義務が設けられておらず、論文等の引用数が米欧に比べて格段に少ないことである。そのため、上述したような分野間の傾向比較などマクロ分析にはむくが、研究成果の個別的な技術波及などミクロな分析には、ほとんど無力となっている。そこで本研究では、サイエンスリンケージの手法が論文等の単純な引用関係を指標とするのに対して、学術文献と技術文献における語彙使用の類似性によって研究内容の類似性を評価する共語分析の手法を導入し、基礎研究—新技術開発の相互連関を語彙レベルで計量することを目標とする。今回の報告では、IT分野を中心にして、科研費に代表される学術研究の成果が、科学技術文献にみられる諸技術にどの程度波及するかを評価し、提案する手法の有用性について考察する。

2. 評価手法の提案

学術研究の成果が科学技術に波及するプロセスを、次のように研究成果報告書に記述されるキーワードの共起の関係として計量する。1) 学術研究の源として文部科学省科学研究費補助金（科研費）採択課題を細目及び年度毎に抽出する。2) 次に、課題の代表申請者を著者として含む科学技術論文を抽出する。3) これら論文に出現するキーワードと科研費成果報告書に記載されたキーワードとの間の類似性を評価し、発表論文との相関を調べる（量的評価）。4) 相関の強い年度につき発表論文のキーワードの共起頻度を算出し、共語関係をマトリクス化する。5) 共語指標として知られている Jaccard 指標及び近接指標により相関を計量し、共語ネットワークとして可視化する（質的評価）。

3. 結果と考察

(1) 科研費データベース KAKEN による細目毎採択課題の抽出

我が国の大学等を中心とした学術研究は、競争的研究資金を基軸に予算が組まれるようになり、その総額の半分を占めるのが科研費である。科研費は、研究者の創意に任せて研究テーマを設定するボトム

表1 電気電子工学分科における 2003 年度終了の科研費採択課題の概況

科研費細目	終了課題数	平均実施期間	平均研究者数	平均 KW 数
5101 電力工学・電気機器工学	73 件	2.34 年	1.99 人	9.07
5102 電子・電気材料工学	84 件	2.34 年	2.04 人	9.40
5103 電子デバイス・電子機器	72 件	2.26 年	2.03 人	8.72
5104 通信・ネットワーク工学	106 件	2.25 年	2.07 人	8.57
5105 システム工学	15 件	2.13 年	1.80 人	8.67
5106 計測工学	57 件	2.25 年	2.21 人	8.63
5107 制御工学	39 件	2.26 年	1.97 人	8.39
51 電気電子工学分科全体	446 件	2.28 年	2.04 人	8.83
科研費全体	18563 件	2.22 年	2.40 人	9.05

アップ型の学術研究制度として発展し、2009 年度においては、採択研数 57000 件（新規及び継続）、総額 1970 億円の規模にある。このような背景から、科研費採択課題は、我が国の学術研究の量と質を反映する中心的な存在とみなすことができる。科研費の採択課題及び成果概要は、国立情報学研究所が公開する KAKEN データベースにアクセスすることにより 1965 年度以降分について知ることができる。今回分析対象としたのは、電気電子工学分科における 7 細目 446 課題であり、2003 年度終了分につきその概況を表 1 に示した。電気電子工学分科を選定した理由は、重点分野とされる IT 分野との関係が強く、サイエンスリンケージとの比較も視野に置いたことによる。

(2) 科研費代表研究者を含む科学技術文献の抽出

代表申請者の氏名及び所属を検索語として、科学技術文献データベース JDreamII を用いて、申請者を著者として含む公表論文を抽出した。発表論文数は、細目によって 40～400 論文／年と差が大きい、図 1 に見られるとおり、概ね 2003 年終了年あたりに発表論文数のピークが見られ、ファンディング年の終了と発表論文数のピークに相関があることが分かった。

(3) データベース間にまたがるキーワードの類似度の算出

一般に、データベース毎にキーワードシソーラスが作成されるために、KAKEN と JDreamII の間においてもキーワードが一致するかどうかは不明である。これらデータベース間のキーワード相関についてはほとんど知見がないため、科研費成果報告書に記載されたキーワードと申請者を含む科学技術文献に現れるキーワードの相関性を検討した。用いた指標は、次式で定義されるコサイン類似度である。

$$\text{コサイン類似度} = \frac{(\vec{A} \cdot \vec{B})}{|\vec{A}| |\vec{B}|}$$

ここで、 $\vec{A}$ 及び $\vec{B}$ は、KAKEN 及び JDreamII のそれぞれのキーワードを次元として定義される文書ベクトルである。図 2 には、論文単位で算出されたコサイン類似度の細目別平均値を年度別に算出した結果を示す。ファンディング終了年を前後して類似度に明確にピークが見出されたことから、終了年に公表された論文は、分野によらず相対的にキーワードの類似度が高いことが分かった。ちなみに、この時の共語数は、全キーワード中 KAKEN では約 15%、JDreamII では約 10%であった。

(4) 共語マップの作成

終了年の発表論文は科研費課題との相関が強いことが明らかになった。次に、申請者を著者として含

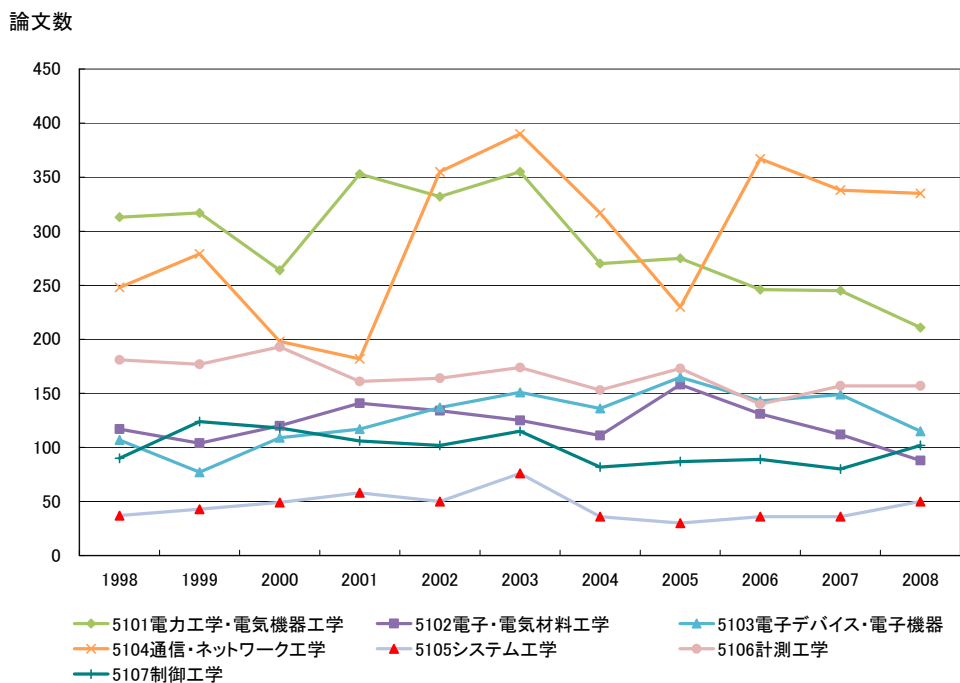


図 1 2003 年度終了課題代表研究者が公表する西暦年ごとの論文数 (JDreamII 和文文献)

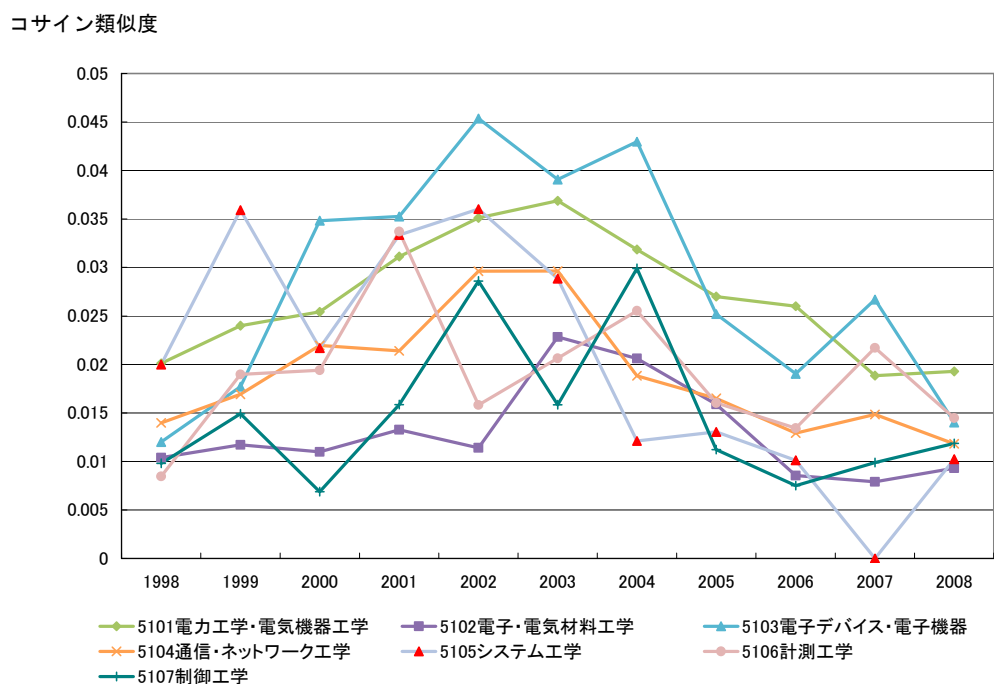


図 2 2003 年度終了課題報告書及び JDreamII 登録文献間での KW のコサイン類似度

む科学技術文献のキーワードが、科研費課題のキーワードとどのような関係にあるかを明らかにするために、キーワードの相関マップを作成した。このようなマップを作成する意義は、学術研究で見いだされたシーズが、どのような科学技術の課題（キーワード）に波及していくかを語彙間の結合によって可視化できることにある。電子・電気材料工学分野について、Jaccard 指標及び近接指標により評価した結果を図 3 及び図 4 に示す。Jaccard 指標は共語マップの中心部（高頻出語）と周辺部（低頻出語）がバランスよく現れる指標であり、近接指標では特に周辺部の用語が強調される⁴。図にみられるとおり

