

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年6月14日現在

機関番号：12601

研究種目：若手研究（B）

研究期間：2009～2011

課題番号：21700266

研究課題名（和文） 引用ネットワーク分析におけるリンク形成手法に関する研究

研究課題名（英文） Link-creating method in citation network analysis

研究代表者

梶川 裕矢（KAJIKAWA YUYA）

東京大学・大学院工学系研究科・特任講師

研究者番号：70401148

研究成果の概要（和文）：

引用ネットワーク分析は、学術分野の全体像を俯瞰するための手法として国内外で広く研究がなされている。その手法として、共引用、書誌結合、直接引用が古くから知られているが、手法間の差異を実証的に検討したものは極めて少ない。本研究では、各手法の有する特徴を定量的に明らかにした。具体的には、特定の分野の論文集合に対し、3種類のリンク形成手法（共引用、書誌結合、直接引用）を用いて引用ネットワークを形成し、それぞれの手法の妥当性、有効性、効率性を評価した。さらに、引用に重みを持たせる手法を提案し分析を行った。その結果、萌芽領域の検出において、直接引用を用いた手法が最も効果的であることが分かった。また、直接引用を用いたネットワークの語の分布の特徴、分野横断性や、分野間の引用の時間遅れを測定する手法に関する基礎的な検討を行った。

研究成果の概要（英文）：

Citation network analysis has been used to comprehend an overview of academic research. Three methods, co-citation, bibliographic coupling, and direct citation, are known, but there are few studies to compare those methods with real data. This research quantitatively revealed characteristics of them. Citation networks were created with those three methods, and evaluated their relevance, effectiveness, and efficiency. A method giving weight on links are proposed and also analyzed. It is found that direct citation is superior to the others in detection of emerging research fields. Term distribution, interdisciplinarity, citation lag were also investigated.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	1,200,000	360,000	1,560,000
2010年度	1,000,000	300,000	1,300,000
2011年度	1,000,000	300,000	1,300,000
年度			
年度			
総計	3,200,000	960,000	4,160,000

研究分野：総合領域

科研費の分科・細目：情報学、図書館情報学・人文社会情報学

キーワード：情報図書館学、情報工学

## 1. 研究開始当初の背景

産業における競争力の源泉が、従来の有形

資産から、知識・技術といった無形資産にシフトするに従い、将来のイノベーションの核

となる技術を連続して産み出すこと、又、新規有望技術分野をいち早く認識し、着手し、製品化まで行うことが、国家、企業、大学、いずれにとっても、競争優位を確保する上で重要な課題となっている。特に、資源の少ない我が国にとって、科学技術創造立国以外に繁栄を維持する道がないのは明白である。しかし、技術体系が細分化・複雑化する中、現在では技術シーズを実用化に結び付ける為の様々な技術の統合が必要となっており、より俯瞰的な視野に立った、戦略的な R&D 戦略の立案、そのための方法論の構築が強く求められている。

特に現在のように、知の総量が爆発的に増大している状況下においては、当該分野の専門家といえどもその全体像を把握することは容易ではない。例えば、DNA に関して出版される論文数は既に年間十万本を超えており、全体像の把握は極めて困難となっている。

引用ネットワーク分析は、学術分野を俯瞰する手法として国内外で広く研究開発が行われており、その成果はこの分野の主要な学術雑誌である *Journal of the American Society for Information Science and Technology (JASIST)* や *Scientometrics* 誌上にて報告されている。良く知られたものとしては、Thomson-ISI 社の Small による SCI-Map、Arizona 大学の Chen による CiteSpace、Oklahoma 州立大学の Morris らによる Time Line Visualization、Sandia 研究所の Boyack らによる VxInsight などがある。国内においても申請者の研究グループ、および、科学技術政策研究所の阪らにより科学マップの作成および手法の開発が行われている。

引用ネットワーク分析の手法としては、従来より、共引用、書誌結合、直接引用の3種がよく知られている。しかし、手法間の差異に注意を払うことがなく、場当たりに手法が選択されているのが現状である。例えば、Small や阪らは共引用、Morris らは書誌結合、申請者らは直接引用を用いている。そのような手法間の差異によって結果に影響が出る事が予想されるため、分析結果の解釈、特に R&D 政策立案等への応用に用いる際には留意が必要と思われるが、手法間の差異に関する実証研究は極めて少ない。少数の例外としては、学術俯瞰のためにジャーナル間の引用ネットワークに着目し、共引用、直接引用を用いた際のジャーナルのカテゴリー類似性を分析した Klavans and Boyack の研究、萌芽的研究領域の検出のために論文間の引用ネットワークに着目し、共引用、書誌結合、直接引用を用いた際の検出感度を分析した申請者らの研究がある。しかし、3種のリンク形成手法を用いた分析結果の妥当性・有効

性・効率性を体系的に検討した研究は存在しない。

## 2. 研究の目的

本研究では、ネットワーク分析におけるリンク形成手法の差異が分析結果にもたらす影響を体系的に整理し実証すること、ならびに、複数のリンク形成手法を組み合わせた新しいリンク形成手法を提案することを目的とし、研究を実施した。

また、上記等で手法の有効性を実証した上で、産業上の重要性が高いバイオやエネルギー分野における萌芽領域の検出や、分野横断性の測定による基盤的研究テーマの抽出、引用の時間遅れに着目した分野間の知識伝播の構造分析を行った。

## 3. 研究の方法

リンク形成手法を評価するために、特定の分野の論文集合に対し、3種類のリンク形成手法（共引用、書誌結合、直接引用）を用いネットワークを構築し、それぞれのネットワークの構造、及び、クラスタリングした結果を分析することで各手法の評価を行った。評価としては、引用ネットワーク分析を当該分野の俯瞰に用いる際の論文の類似度から見た妥当性、萌芽的分野の検出に用いる際の有効性に着目し、評価を行った。

また、手法の頑健性を評価するために、クラスタリング手法に関する研究を行った。エネルギー分野、環境分野、応用物理・材料分野等を対象に、引用ネットワークを形成し、クラスタリングの精度を比較した。

## 4. 研究成果

表1は、複数のリンク形成手法を萌芽的な分野の検出の容易さ(Visibility)、内容の適切性(Density)、抽出された領域の萌芽性の度合い(Speed)で評価した結果の例である。表1から直接引用が最も容易に、適切に、早く萌芽領域を抽出できることが分かった。

表2はリンクに対する重みづけの効果を検討した結果である。重みの種類は、(i) No weight, (ii) Frequency of citations, (iii) Difference of publication years, (iv) Reference Similarity, (v) Keyword Similarity であり、Reference Similarity を用いたものが萌芽領域の検出に有効であることが分かった。

また、分野横断的な研究領域の検出においては、個々の論文の引用のネットワーク内での分散に注目したモデル化を行った。その結果、引用が多くの領域に分散している論文は、様々な領域に応用可能な概念や手法の分析等の論文が多いことが分かった(図1)。

また分析結果の頑健性を検討するために、異なるサイズのコーパスに対するクラスタ

リングの品質を評価した結果、引用ネットワークに対し、クラスタリングを用い頑健な結果を得るためには、分野にかかわらず、約百～数千の論文数が必要なことが分かった(図2)。この結果は、クラスタリングを引用ネットワークに対し応用する際に極めて有用なガイドラインとなる。

表1 リンク形成手法の比較

Year	Direct Citation			Co-Citation			Bibliographic coupling		
	Size	Density	Avg. year	Size	Density	Avg. year	Size	Density	Avg. year
<b>(i) No weight</b>									
2000	24	1.310	1999.57	0	0.000	0.00	0	0.000	0.00
2001	12	0.392	2000.37	0	0.000	0.00	0	0.000	0.00
2002	12	0.325	2001.26	0	0.000	0.00	19	10.650	2001.42
2003	12	0.139	2001.90	0	0.000	0.00	27	5.530	2001.93
2004	17	0.067	2002.74	26	0.385	2001.19	18	8.496	2002.93
2005	16	0.061	2003.60	29	0.458	2002.18	34	2.750	2003.34
2006	17	0.049	2004.40	33	0.429	2002.88	33	3.340	2004.14
2007	17	0.041	2005.13	26	0.550	2003.95	20	10.422	2005.23
<b>(ii) Frequency of citations</b>									
2000	24	1.310	1999.57	0	0.000	0.00	0	0.000	0.00
2001	12	0.392	2000.37	0	0.000	0.00	0	0.000	0.00
2002	12	0.325	2001.26	0	0.000	0.00	19	10.650	2001.42
2003	12	0.139	2001.90	0	0.000	0.00	27	5.530	2001.93
2004	17	0.067	2002.74	34	0.426	2001.21	18	8.496	2002.93
2005	16	0.061	2003.60	33	0.363	2002.07	34	2.750	2003.34
2006	17	0.049	2004.40	28	0.512	2002.39	33	3.340	2004.14
2007	17	0.041	2005.13	26	0.495	2003.38	20	10.422	2005.23
<b>(iii) Difference of publication years</b>									
2000	17	0.996	1999.45	0	0.000	0.00	0	0.000	0.00
2001	8	0.621	2000.48	0	0.000	0.00	0	0.000	0.00
2002	11	0.391	2001.32	0	0.000	0.00	0	0.000	0.00
2003	9	0.221	2001.96	0	0.000	0.00	11	5.358	2000.60
2004	10	0.131	2002.81	26	0.629	2001.23	14	6.322	2001.47
2005	11	0.092	2003.61	29	0.455	2002.13	13	4.865	2002.11
2006	15	0.055	2004.43	23	0.750	2002.56	8	4.969	2003.61
2007	13	0.053	2005.10	22	0.719	2003.79	10	4.370	2005.91
<b>(iv) Reference similarity</b>									
2000	16	1.333	1999.81	0	0.000	0.00	0	0.000	0.00
2001	8	0.638	2000.47	0	0.000	0.00	0	0.000	0.00
2002	11	0.405	2001.38	0	0.000	0.00	0	0.000	0.00
2003	9	0.220	2002.01	0	0.000	0.00	19	10.784	2001.39
2004	11	0.131	2002.89	21	0.910	2001.79	17	12.154	2002.08
2005	11	0.091	2003.63	23	0.625	2002.48	15	11.423	2002.95
2006	12	0.068	2004.41	14	1.570	2003.37	16	11.369	2003.71
2007	13	0.056	2005.12	16	1.094	2004.11	18	10.904	2004.48
<b>(v) Keyword similarity</b>									
2000	16	1.421	1999.80	0	0.000	0.00	0	0.000	0.00
2001	11	0.465	2000.35	0	0.000	0.00	0	0.000	0.00
2002	12	0.359	2001.28	0	0.000	0.00	0	0.000	0.00
2003	12	0.146	2001.93	0	0.000	0.00	18	11.464	2001.43
2004	15	0.085	2002.84	21	0.854	2001.78	18	11.753	2002.10
2005	15	0.067	2003.62	29	0.458	2002.18	20	7.506	2002.98
2006	16	0.050	2004.39	24	0.725	2003.25	17	10.619	2003.72
2007	15	0.048	2005.11	25	0.588	2004.03	18	10.573	2004.48

表2 リンクの重みづけの効果

	Visibility (normalized size)
Direct citation	(iv) > (i) = (ii) > (v) > (iii)
Co-citation	(iv) > (ii) > (v) > (i) > (iii)
Bibliographic citation	(iv) > (ii) > (i) > (v) > (iii)
	Topological relevance (density)
Direct citation	(iv) > (v) > (iii) > (i) = (ii)
Co-citation	(iv) > (ii) > (v) > (iii) > (i)
Bibliographic citation	(iii) > (iv) > (v) > (ii) > (i)
	Speed (average birth year)
Direct citation	(iii) > (i) = (ii) = (iv) = (v)
Co-citation	(iii) > (ii) > (i) = (iv) = (v)
Bibliographic citation	(iii) > (ii) > (i) = (iv) = (v)

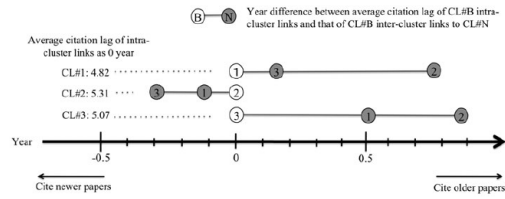


図1 分野横断性の測定

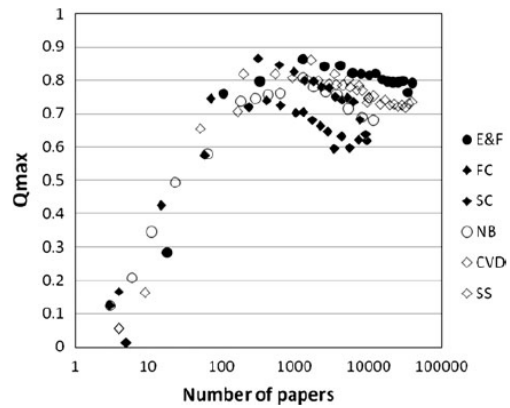


図2 クラスタリングの頑健性

分野横断性に関する研究では、サステイナビリティ学を対象に、複数の論文を繋いでいる論文の抽出を行った。媒介中心性やハーフィンダル指数等の複数の指標を用い、各論文に対しスコアを付与し、その論文の有する学際性の評価を行った。その結果、複数の指標を組み合わせ使用することにより、汎用性の高いツールの開発や、分野横断性の高い政策、哲学的考察等、複数の異なる特徴を有する論文を選択的に抽出可能であることが分かった。この結果を用いることで、サービスサイエンス等の分野横断性の高い他の領域においても、学際性の高い有用な論文を抽出することが可能であることが分かった。

現在のように知の総量が爆発的に増大している状況下において、政策・戦略立案者が、特定の学術分野の俯瞰を行うこと、特に、現在現れつつある萌芽的な研究領域の検出をいち早く行うこと、そしてそれらをサポートするツールへの需要は極めて大きい。本研究では、直接引用を用いた引用ネットワーク分析により、グリーンイノベーション分野やサービスイノベーション分野など、社会的な関心が高く、研究領域としても萌芽期～成熟期に当たるものを選択し、学術分野の俯瞰を行うとともに、その分野において急速に伸びている領域の同定を行った。その結果、提案手法を用いて、当該学術分野の俯瞰、萌芽的な研究領域の検出などを行うことが可能となった。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 6 件)

Y. Takeda and Y. Kajikawa "Tracking modularity in citation networks: a comparative study on clustering quality of citation network" *Scientometrics* 83 (2010) 783-792.

Y. Kajikawa, Y. Takeda, and K. Matsushima "Computer-assisted roadmapping: A case study in energy research" *Foresight* 12 (2010) 4-15.

梶川裕矢、「リンクマイニングを用いた引用情報の活用」、*情報の科学と技術* 60 (2010) 224-229.

N. Shibata, Y. Kajikawa, and I. Sakata, "Measuring relatedness between communities in a citation network", *Journal of the American Society for Information Science and Technology* 62 (2011) 1360-1369.

N. Shibata, Y. Kajikawa, Y. Takeda, I. Sakata, and K. Matsushima, "Detecting emerging research fronts in regenerative medicine by the citation network analysis of scientific publications", *Technological Forecasting and Social Change* 78 (2011) 274-282.

H. Nakamura, S. Suzuki, H. Tomobe, Y. Kajikawa, and I. Sakata, "Citation lag analysis in supply chain research", *Scientometrics* 87 (2011) 221-232.

[学会発表] (計 4 件)

Y. Kajikawa and J. Mori, "Interdisciplinary Research Detection by Citation Indicators", *IEEE International Conference on Industrial Engineering and Engineering Management 2009 (IEEM2009)* in Hong Kong. (December 8-11, 2009).

Y. Kajikawa, and J. Mori, "Interdisciplinary research detection in Sustainability Science", *Tracking and evaluating interdisciplinary research: metrics and maps, workshop in conjunction with 12th International Conference on Scientometrics and Informetrics (ISSI2009)*, in Rio de Janeiro, Brazil (14-17, July, 2009).

Y. Kajikawa, H. Kashima, J. Mori, and N. Shibata, "Technological forecasting by citation network analysis", *2010 Annual Meeting of Institute for Operations Research and*

*Management Sciences (INFORMS2010)* in Austin. (November 7-10, 2010).

Y. Kajikawa, "Text and link mining: New tools for technology and innovation management", *Portland International Conference on Management Engineering and Technology 2011 (PICMET'11)* in Portland, U.S.A. (July 31 - August 4, 2011).

## 6. 研究組織

(1) 研究代表者

梶川 裕矢 (KAJIKAWA YUYA)

東京大学・大学院工学系研究科・特任講師  
研究者番号：70401148