

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 11 日現在

機関番号：12601

研究種目：若手研究(B)

研究期間：2012～2013

課題番号：24700246

研究課題名(和文)イノベーションに資する学術研究領域の特定と知識伝搬に関する研究

研究課題名(英文)Detection of Promising Science Fields and Knowledge Transition among the Heterogeneous Networks.

研究代表者

佐々木 一 (Sasaki, Hajime)

東京大学・政策ビジョン研究センター・特任研究員

研究者番号：40584199

交付決定額(研究期間全体)：(直接経費) 1,400,000円、(間接経費) 420,000円

研究成果の概要(和文)：将来の産業の発展に資する科学領域を早期に発見し集中的に投資することは、わが国のイノベーション戦略において重要課題の一つである。科学知識と産業技術が密接に関係することは、ビジネスにおいてはもとより、研究活動においてもポジティブな影響を与えるとされている。学術論文および特許公報の書誌情報に対して時空間ネットワークの適用をし、異なるネットワークドメイン間における知識遷移を捉えることで、科学技術イノベーションの早期発見に新たな視座を与えられると考えた。太陽光発電領域について、学術論文と特許公報の書誌情報を用いて、時空間ネットワークと最大流量問題を適用し科学から技術への知識遷移について論じた。

研究成果の概要(英文)：Science and Technology oriented countries have to invest intensively in science fields that will contribute as Innovation in future. The purpose of this research is to suggest a methodology for identifying promising fields to grasp dynamic transitions between science and technology and graphically representing the relations of information among papers or patents linked by citations. In this paper, we focused on the photovoltaic field as a case study. We discussed knowledge transition from science to technology by time expansion network analysis between academic papers and patent publications.

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：イノベーション 有望領域 早期特定 異種ネットワーク 書誌情報分析 太陽光発電技術 知識伝搬
サイエンスリンケージ

1. 研究開始当初の背景

将来の産業の発展に資する科学領域を早期に発見し集中的に投資することは、わが国のイノベーション戦略において重要課題の一つである。科学知識と産業技術が密接に関係することは、ビジネスにおいてはもとより、研究活動においてもポジティブな影響を与えるとされている(Zucker and Darby, 2000; Hall et al., 2001; Azoulay et al., 2006)。(Rosenberg, 1998)によれば、科学知識と産業技術が密接な関係を持つことで、新たなアイデアが創出されることが分かっている。また、十分なリソースを得られず水面下で実施されてきたような研究活動が、民間企業を通じて結実しうることも事実である(Agrawal and Henderson, 2002)。(Narin et al., 1997)によれば、米国特許のうち科学研究論文を引用している特許の割合は、1987年から1994年の間で急激に増加している。このことは、科学知識が産業技術に与える影響がますます強くなってきていることを示唆するものである。

科学知を記述した学術論文の量は、近年爆発的な増加の一途をたどっている。多くの学術分野では、今や年間に出版される論文数は数千から数万件以上となっており、当該分野の専門家といえども全体像を把握することは容易ではない。一方で、産業知の中でも新規性の有る先端技術を記述した特許公報についても同様の現象が起こっている。現在、そのように膨大な科学知や技術知から将来有望な領域を特定することが課題となっている。特に、科学が技術に至る知識の遷移を特定することが可能となれば、将来の科学技術イノベーションに資する領域の早期特定に資する。しかしながら特定領域について、科学から技術への知識の変遷を捉えることで、将来の産業の発展に資する可能性がある学術領域を特定する手法については十分に議論はなされていない。

2. 研究の目的

本研究の目的は、将来の産業の発展に大きく資する可能性が高い学術領域を、情報学的手法を用いて膨大な数の文献との中から、早期に特定する手法を提案することである。

本研究では過去の事例をもとに特定の科学技術知識が、科学の層から技術の層へと伝搬して行く過程をダイナミクスで捉える。学術論文と特許公報の間の知識遷移および学術論文の層におけるダイナミクスの中から学術分野のネットワーク特徴量の推移の性格を特定し、将来における新規有望技術領域・論文の予測を行うことを目的とした。

(Shibata et al, 2010)は太陽光発電技術に関する学術論文と特許論文の領域を比較し、研究開発戦略における機会発見に資する研究を行なった。これは、特定の時期におけるスナップショットでの議論を行ったものであり、時間軸を伴う相互の知識遷移について

は十分に議論がなされていない。(Small H., 2007; 七丈, 2013)は、論文クラスタ間の継承関係の時系列的变化を動的に捉えることによる萌芽的領域の特定に資する手法を提案しているが、いずれも科学と技術の関係について十分議論されたものではない。筆者は、学術論文および特許公報の書誌情報に対して時空間ネットワークの適用をし、異なるネットワークドメイン間における知識遷移を捉えることで、科学技術イノベーションの早期発見に新たな視座を与えられると考えた。時空間ネットワーク(Time Expanded Network)は、ネットワークを時間軸に発展させたものであり一般的なネットワーク問題に対して時間発展要素を導入したものである。近年時空間ネットワークを用いた応用研究が複数されている(F.G. Engineer et al., 2011; N. Shah et al., 2012)が、書誌情報を用いた科学技術の知識遷移については十分に議論がなされていない。本研究では、学術論文および特許公報のそれぞれのレイヤにおける時間的变化の様子を、各出版年時におけるクラスタをメタノードとした引用関係の重み付き時空間ネットワークとして扱う。また、学術論文と特許公報の間の関係は、それぞれのクラスタが持つ特徴語の類似度に基づいた重み付き時空間ネットワークとして扱う。これにより、対象分野の研究から技術への知識遷移について総合的に俯瞰できると考える。

3. 研究の方法

本研究では、分析のケーススタディとして、太陽光発電に関する学術論文と特許公報を分析対象とした。これらの文書の抄録文字列、出版(出願)年、直接引用関係をデータセットとする。2002年から2012年に出版された学術論文および、出願された特許公報を対象に、年次ごとに引用ネットワークを構築した。ここで得られた全ての年次別ネットワークに対し、クラスタリングを行う。クラスタリングにはNewman法(Newman et al, 2004)を用いる。Newman法はネットワークにおけるモジュラリティ Q (式1)が最大となるエッジを見つけ、そのエッジをクラスタの断面として見なすものである。

$$Q = \sum_i (e_{ii} - a_i^2) \quad \dots \text{式1}$$

ここで、 e_{ii} は同じクラスタ g_i に属するノード同士が繋がるリンク数の全リンク数に占める割合であり、 a_i はリンクの一方に g_i 内のノードの「手」が選ばれる確率である。

学術論文、特許公報それぞれについて各年次上位5クラスタをメタノードとし、新たにクラスタ遷移を表す時空間ネットワークを作成する。その際、各書誌内におけるクラスタ間の引用情報を、メタノード間を繋ぐエッジの重みとすることで、重み付き時空間ネットワークとして見なす。

一方、得られた各クラスタに属する書誌から

抽出した抄録文字列を元に、年次別クラスタごとの特徴語を抽出する。特徴語の抽出は、一般的に用いられる *tf-idf* において、文書頻度をクラスタ頻度に置き換えた指標 (式 2) を用いる。

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right) \quad \dots \text{式 2}$$

ここで、 $tf_{i,j}$ はクラスタ j の中に出現する単語 i の数、 df_i は、単語 i を含むクラスタの数、 N はクラスタの数を表す。

特徴語指標を元にコサイン類似度 (式 3) を算出し、類似度行列を作成する。

$$sim_{tfidf}(s, t) = \text{Cosine}(s, t) = \vec{w}_s \cdot \vec{w}_t = \sum_i w_s^{(i)} w_t^{(i)}$$

・・・式 3

得られた学術論文-特許公報間類似度行列から、同年次のクラスタ同士に渡るエッジは除去した上で、類似度を重みとした時空間ネットワークを作成する。同年次同士に渡るエッジを除去した理由は、学術論文と特許技術の間における知識の遷移をモデル化するため、同年次内での知識遷移は考えにくいと判断したことによる。

得られた特徴語類似度ネットワークを元に、特定のノード同士の最大流量アルゴリズムに基づくクラスタリングを行う。最大流量アルゴリズムでは、ネットワークにおけるエッジの重みを、そのエッジが流すことの出来る最大の流量としてみなす。任意の Source ノードから Sink ノードに至る流れの経路のうち最大の流量を流すことが出来る経路を特定することを目的としたアルゴリズムである。最大流量クラスタリングは最大流量最小カットの定理にもとづき、ネットワークコミュニティの分割に適用する手法である。具体的には (Ino, H et al., 2005) による Web コンテンツのコミュニティ抽出などの応用例があげられる。また、(Horiike et al., 2009; Victor, S et al., 2012) は、研究者ネットワークにおけるコミュニティ抽出として、最大流量クラスタリングを活用している。本研究では、(A. V. Goldberg and R. E. Tarjan, 1988) によって提案されたプッシュリラベル法 (P/R 法) を適用し、論文と特許の間の流量を元に、科学から技術への知識遷移を特定する。

4. 研究成果

論文データベース Web of science より、“photovoltaic” および “solar cell” をクエリとして学術論文を抽出した結果 50,913 件の学術論文を得た。また、特許データベース Thomson Innovation から同じ検索クエリにより特許公報を抽出した結果、63,972 件の特許公報を得た。

表 1 は 2012 年までに出版された学術論文および特許公報それぞれの上位 5 クラスタにおける特徴語を元に推測した、領域の内容をラベリングしたものである。

表 1 クラスタ別領域内容ラベル

	学術論文	特許公報
第1クラスタ	C1 Thin film	Conductor
第2クラスタ	C2 Organic	Protection sheet of for panel
第3クラスタ	C3 Dye sensitized	Dye sensitized
第4クラスタ	C4 Power, point tracking	Multi junction
第5クラスタ	C5 HgCdTe photovoltaic	Silicon semiconductor

学術論文における第 1 クラスタは薄膜系太陽光発電に関する研究が多いことを意味し、第 2 クラスタは有機系、第 3 クラスタに色素増感系太陽光発電を表す。第 4 クラスタは太陽光発電における出力を最大に出来る最適な Point tracking について論じているクラスタであることがわかった。また、第 5 クラスタはカドミウムテルル太陽光発電に関する研究が大勢を占めると評価したことを意味する。一方で特許公報では、第 1 クラスタは電極導体に関する技術が多くを占めており、第 2 クラスタとして太陽光パネルにおけるカバーガラス等の保護技術について議論したものが多。第 3 クラスタは学術論文と同じく、色素増感系太陽光発電技術に関する特許が多いことが分かった。第 4 クラスタでは多接合型の太陽光発電技術、第 5 クラスタではシリコン系太陽光発電技術に関する特許が比較的多いことがわかった。

次に、学術論文および特許公報の各年上位 5 クラスタ同士の特徴語に対して、クラスタ間のコサイン類似度を算出した。図 1 はその結果に基づくヒートマップである。

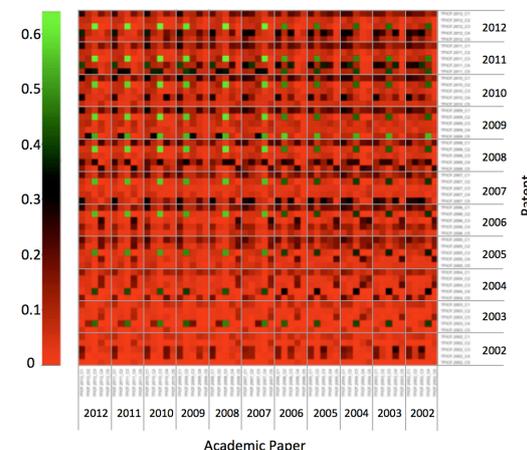


図 1 論文特許間類似度推移

表 1 で得られた結果と同様、学術論文における 2012 年の第 3 クラスタと特許公報における 2012 年の第 3 クラスタが最も類似度が高い。同程度の類似度を有するクラスタのペアは 2005 年程度まで遡ることが見て取れる。クラスタの時系列における現象は、発生 (born)、消滅 (vanish)、成長 (grow)、衰退 (shrink)、結合 (merge)、分轄 (split) といった形態が存在する (Péter E et al., 2013)。したがって、各年で特定されたクラスタは、同じ層のネットワークであっても、一定程度の新陳代謝が存在するものであり、常に同一のものではない。

特許公報の 2012 年時点における第 3 クラスタをさかのぼり、どの時点で研究領域からの遷移を得られたかを捉えるため、類似度ネットワークに対して、最大流量計算を行なった。Source ノードは特許公報における 2012 年の第 3 クラスタとし、Sink ノードの選定は、2002 年の学術論文クラスタ 5 つ全てに対して最大流量計算を行ない、最大の流量を得られた 2002 年の第 4 クラスタとした。

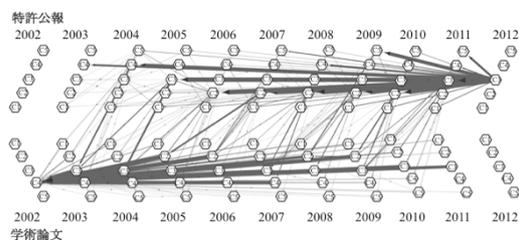


図 2 論文-特許間最大流量ネットワーク

図 2 は、最大流量を重みとしたネットワークである。なお、sink ノードと直接学術論文ネットワークに渡るエッジ、および source ノードと直接特許公報ネットワークに渡るエッジは除去してある。学術論文ネットワークと特許公報ネットワークの間で最も流量が大きいエッジとして、学術論文 2005 年第 2 クラスタと特許公報 2007 年第 2 クラスタ間のエッジがあげられる。特徴語の類似度ではこの量クラスタ間に特段大きな類似度は確認出来ないにも関わらず、複数ある論文特許間のエッジの中で最も大きな流量が流れるエッジである。相互のクラスタの特徴語について上位 10 語をリストし表 2 に示す。いずれのクラスタにおいても色素増感系太陽電池を表す“dye sensitized”が複数確認できる。

表 2 特徴語上位 10 語

学術論文(2005. C2)	特許公報(2007. C2)
tio2	dye
dye	sensitized
polymer	dye sensitized
sensitized	dye sensitized solar
dye sensitized	dye sensitized solar cell
sensitized solar	sensitized solar cell
sensitized solar cell	sensitized solar
nanocrystalline	sensitizing
dye sensitized solar	pigment
dye sensitized solar cell	electrode

次に、太陽光発電の学術領域において、クラスタのネットワーク特徴量の推移を算出した。ここでは、クラスタ間の引用関係に基づく時系列ネットワークを元に、各年次のクラスタがどのように推移してきたかを特定する。図 3 では各年次のクラスタの次数中心性がその後どのように遷移してきたかを表した。2004 年までの年次においては、第 3 クラスタの次数中心性がその後の推移を経て

2012 年時点でもっとも高くなるが、2005 年以降では第 2 クラスタが 2012 年時点の次数中心性が高いことが確認できた。

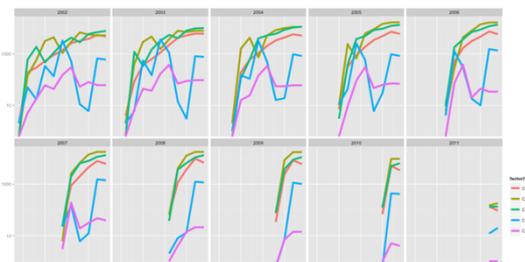


図 3 各年次クラスタの次数中心性推移

本研究では、学術論文と特許公報の書誌情報を用いて、時空間ネットワークと最大流量問題を適用し科学と技術の知識遷移について論じた。将来における新規有望技術領域・論文の予測に資するモデル化については達成できなかったものの、新たな視点を提案したことによる一定の洞察は得られた。今後他の領域についても拡張していくことで、新規有望技術領域の予測研究を発展・実用化させていく必要がある。

参考文献

Zucker, L.G., Darby, M.R., “Capturing technological opportunities via Japan’s star scientists.”, *Journal of Technology Transfer* 26, 37-58, 2000.

Hall B.H., Link A.N., Scott J.T., “Barriers inhibiting industry from partnering with universities: evidence from the advanced technology program.”, *Journal of Technology Transfer* 26, 87-98, 2001.

Azoulay P., Ding W., Stuart T., 2006. “The impact of academic patenting on the rate, quality and direction of (public) research.”, NBER working paper 11917, Cambridge, MA., 2006.

Rosenberg N., “Chemical engineering as a general purpose technology. In: Helpman, E. (Ed.), *General Purpose Technologies and Economic Growth*.”, MIT Press, Cambridge, pp. 167-192, 1998.

Francis Narin, Kimberly S. Hamilton and Dominic Olivastro, “The increasing linkage between U.S. technology and public science”, *Research Policy*, Volume 26, Issue 3, October 1997, Pages 317-330, 1997.

N. Shibata, Y. Kajikawa, and I. Sakata. "Extracting the commercialization gap between science and technology - case study of a solar cell.", *Technological Forecasting and Social Change*, 77:1147-1155, 2010.

Small H., "Citation Structure of an Emerging Research Area: Organic Thin Film.", *Proceedings of ISSI*, pp. 718-725. 2007.

七丈, "共引用クラスタリングによる研究分野の動的把握に向けた試論", *情報知識学会誌* 2013, Vol.23, No.3, 371-379, 2013.

F.G. Engineer, G.L. Nemhauser, and M.W.P. Savelsgergh, "Dynamic programming-based column generation on Time-Expanded Network: Application to the Dial-a-Flight problem", *INFORMS Journal on Computing*, 23(1), pp. 105-119, 2011.

N. Shah, S. Kumar, F. Bastani, and I.L. Yen, "Optimization models for assessing the peak capacity utilization of intelligent transportation systems", *European Journal of Operational Research*, 216, pp. 239-251, 2012.

Newman M.E.J., "Fast algorithm for detecting community structure in networks", *Physical Review E*, Vol. 69, p. 066133, 2004.

Ino, H, Kudo. M, Nakamura. A, "A Comparative Study of Algorithms for Finding Web Communities", *Data Engineering Workshops*, 2005. 21st International Conference, 1257, 2005.

Horiike. T, Takahashi. Y, Kuboyama. T and Sakamoto. H, "Extracting Research Communities by Improved Maximum Flow Algorithm", *Knowledge-Based and Intelligent Information and Engineering Systems Lecture Notes in Computer Science* Volume 5712, pp472-479, 2009.

Victor Ströele, Geraldo Zimbrão, Jano M. Souza, "Modeling, Mining and Analysis of Multi-Relational Scientific Social Network", *Journal of Universal Computer Science*, vol. 18, no. 8 (2012), 1048-1068, 2012.

A. V. Goldberg and R. E. Tarjan, "A New Approach to the Maximum Flow Problem", *Journal of the ACM* 35:921-940, 1988.

Péter Érdi, Kinga Makovi, Zoltán Somogyvári, Katherine Strandburg, Jan Tobochnik, Péter Volf, László Zalányi, "Prediction of Emerging Technologies Based on Analysis of the U.S. Patent Citation Network", *Scientometrics: Volume 95, Issue 1 (2013)*, Page 225-242, 2013.

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)

[学会発表] (計1件)

1. H. Sasaki, Y Kajikawa, I. Sakata, V. Ittipanuvat, Predicting the Potential Industrial Fields of Technological Spin-offs by Using IPC in Patent Analysis, *Technology Management for Emerging Technologies (PICMET)*, 2012 Proceedings of PICMET '12, July 29 2012-Aug. 2, pp.975 - 991, 2012.

6. 研究組織

(1) 研究代表者

佐々木 一(SASAKI Hajime)

東京大学・政策ビジョン研究センター・特任研究員

研究者番号 : 40584199