

AIにおけるサイエンスとイノベーションの共起化：
米国における論文・特許データベースを用いた分析

Co-occurrence of Science and Innovation in AI:
Empirical Analysis of Paper-patent Linked
Dataset in the United States

2018年7月

文部科学省 科学技術・学術政策研究所

第1研究グループ

元橋 一之

本 DISCUSSION PAPER は、所内での討論に用いるとともに、関係の方々からの御意見を頂くことを目的に作成したものである。

また、本 DISCUSSION PAPER の内容は、執筆者の見解に基づいてまとめられたものであり、必ずしも機関の公式の見解を示すものではないことに留意されたい。

The DISCUSSION PAPER series is published for discussion within the National Institute of Science and Technology Policy (NISTEP) as well as receiving comments from the community.

It should be noticed that the opinions in this DISCUSSION PAPER are the sole responsibility of the author(s) and do not necessarily reflect the official views of NISTEP.

【執筆者】

元橋 一之

東京大学大学院工学系研究科 教授
文部科学省科学技術・学術政策研究所 客員研究官
独立行政法人経済産業研究所 ファカルティフェロー

【Authors】

Kazuyuki Motohashi Professor, Graduate School of Engineering, The University of Tokyo
Affiliated Fellow, National Institute of Science and Technology Policy (NISTEP), MEXT
Faculty Fellow, Research Institute of Economy, Trade and Industry (RIETI)

本報告書の引用を行う際には、以下を参考に出典を明記願います。
Please specify reference as the following example when citing this paper.

元橋一之 (2018) 「AI におけるサイエンスとイノベーションの共起化：米国における論文・特許データベースを用いた分析」, *NISTEP DISCUSSION PAPER*, No.160, 文部科学省科学技術・学術政策研究所.

DOI: <http://doi.org/10.15108/dp160>

Kazuyuki Motohashi (2018) “Co-occurrence of Science and Innovation in AI: Empirical Analysis of Paper-patent Linked Dataset in the United States,” *NISTEP DISCUSSION PAPER*, No.160, National Institute of Science and Technology Policy, Tokyo.

DOI: <http://doi.org/10.15108/dp160>

AIにおけるサイエンスとイノベーションの共起化： 米国における論文・特許データベースを用いた分析

文部科学省 科学技術・学術政策研究所 第1研究グループ

要旨

本研究においては、米国の企業・研究機関に所属する研究者について、科学技術論文と特許データを著者・発明者レベルで接続を行い、研究者単位でみたサイエンスリンケージのトレンドについて分析を行った。全体的な論文数が増加する一方で、特許発明者による論文数の割合は低下している。一方で、特許からみたサイエンスリンケージ(論文著者による特許発明の割合)は増加傾向にあることが分かった。またAI分野にフォーカスした詳細分析によると、やはり企業著者による論文数シェアは低下傾向にあるが、特許発明者による論文数については企業シェアの低下は見られていない。企業セクターにおいては、オープンに公表される科学技術論文にも取り組みながら、同時に特許による技術の囲い込みを行っており、その傾向が高まっていることが明らかになった。一方で、大学等の公的研究セクターにおいても、論文著者が特許活動にも乗り出す傾向にあり、オープンなサイエンスと特定の所有者による技術の権利化が同時進行で進んでいる姿が浮き上がった。

Co-occurrence of Science and Innovation in AI: Empirical Analysis of Paper-patent Linked Dataset in the United States

First Theory-Oriented Research Group, National Institute of Science and Technology Policy (NISTEP), MEXT

ABSTRACT

In this paper, the trends of science linkage of innovation are analyzed, based on the linked dataset of research articles and patent information for the researchers affiliated with US organizations. The number of overall publication increases, while the share of publications by those who are also patent inventor decreases. On the other hand, the science linkage on the patent side, i.e. the share of patents by those who are authors of research articles increases. When we focus on the AI field, the share of publication by private firms decreases again, but the firm authors share does not decrease for the publications by the author/inventor researchers. A private firm actively engages in patenting activities to protect their propriety technologies. The author at a non-profit research organization, such as university and government laboratory, not only publishes but also patents. In conclusion, both open science and propriety patent are progressed parallelly to push the technology frontier of AI.

[空白のページ]

AI におけるサイエンスとイノベーションの共起化：米国における論文・特許データベースを用いた分析¹

Co-occurrence of science and innovation in AI: Empirical analysis of paper-patent linked dataset in the United States

元橋一之（東京大学、文部科学省科学技術・学術政策研究所、経済産業研究所）

要旨

本研究においては、米国の企業・研究機関に所属する研究者について、科学技術論文と特許データを著者・発明者レベルで接続を行い、研究者単位でみたサイエンスリンケージのトレンドについて分析を行った。全体的な論文数が増加する一方で、特許発明者による論文数の割合は低下している。一方で、特許からみたサイエンスリンケージ（論文著者による特許発明の割合）は増加傾向にあることが分かった。また AI 分野にフォーカスした詳細分析によると、やはり企業著者による論文数シェアは低下傾向にあるが、特許発明者による論文数については企業シェアの低下は見られていない。企業セクターにおいては、オープンに公表される科学技術論文にも取り組みながら、同時に特許による技術の囲い込みを行っており、その傾向が高まっていることが明らかになった。一方で、大学等の公的研究セクターにおいても、論文著者が特許活動にも乗り出す傾向にあり、オープンなサイエンスと特定の所有者による技術の権利化が同時進行で進んでいる姿が浮き上がった。

キーワード：サイエンスリンケージ、論文データ、特許データ、人工知能

¹ 本論文は科学技術・学術政策研究所(NISTEP)におけるイノベーションプロセスデータベースの開発に関する研究成果である。論文作成にあたってデータ作成や助言を頂いた同研究所の塚田主任研究官及び経済産業研究所池内研究員に感謝の意を表したい。なお、本論文における見解は著者個人のものであって、著者が所属する機関のものではない。

1. はじめに

大学や公的研究機関における科学的知見（サイエンス）と産業におけるイノベーションの関係については、さまざまな分析が行われてきている。特に特許や論文といった研究成果や発明単位の分析については、特許の非特許文献（その多くが科学論文）に関する情報を用いることが一般的である（Narin and Noma, 1985; Schmoch 1997）。しかし、この方法は科学的知見→イノベーションというリニアモデルを前提としており、サイエンスとイノベーションがともに進む共起化の状況を示す指標とはなっていない。一方で、従来からサイエンス型産業といわれてきた医薬品産業のみならず、最近では IT やナノテクの進歩に伴い産業界のイノベーションにおける科学的知識の重要性は高まっている（元橋、2014）。このような問題意識に従って、著者らの科学技術・学術政策研究所と経済産業研究所の共同プロジェクトとして、学術論文の著者と特許の発明者を接続することで、サイエンスとイノベーションの共起化を図る新しい指標の開発に取り組んでいる（Ikeuchi et al., 2017）。

ここでは、米国の企業、研究機関の所属する研究者について、著者、発明者の接続を行い、学術論文でみるサイエンスと特許で見るイノベーションの相互関係について分析を行った。また、AI（人工知能）の分野を取り上げて、より詳細な分析を行った。コンピュータ能力の向上とビッグデータの出現によって、ディープラーニングをはじめとする機械学習の新しい手法の開発が急速に進んでいる。データの種類の応じた新しいモデル（例えば画像データに対する CNN、自然言語処理に対する RNN など）や効率的な機械学習の方法（強化学習や GAN など）については、アカデミックな研究成果として公表されている。その一方でこれらの科学的知見のビジネスへの実装も進んでおり、特許出願も進んでいる。特にこの分野においてはグーグルなどの企業研究者による論文での研究成果の公表も進んでおり、サイエンスとイノベーションの共起化が進んでいると考えられる。

本論文の構成は以下のとおりである。まず次節においては、データベースの構築の手法とその結果について述べる。第 3 章においては構築されたデータベースを用いてサイエンスリンケージのトレンドについて明らかにする。第 4 章 AI 分野にフォーカスして、当該分野におけるサイエンスとイノベーションの共起化の状況について詳細な分析を行う。最後に第 5 章においてまとめと今後の研究課題について述べる。

2. 論文データと特許データの接続

論文データとしては Elsevier 社の SCOPUS データを用いて、著者の所属先住所が米国であるレコードを抽出した（NISTEP における 2016 年時点の分析用 MySQL データ）。特許に関しては米国の研究者を対象としたデータ構築を行うため、米国特許商標庁（USPTO）のデータを用いた。具体的には、2017 年 4 月 18 日時点で同機関のデータベースサイトである patentview.org からダウンロードしたデータを用いた。

なお、論文については公表年が 1996 年～2014 年（若干の 2015 年データ）約 860 万件、特許については約 300 万件（ただし、出願年が 1996 年以降のものは約 200 万件）を対象と

している。ただし、特許データについては、権利化された特許の情報のみが公開されているため、特許審査期間のラグによって最新の出願年は 2011 年となっている。

論文著者と特許発明者の接続については、日本の研究者に関する先行研究 (Ikeuchi et al., 2017) にならい、著者・発明者の氏名と所属情報を用いて行った。まず、接続を行うためにそれぞれのデータの所属先のタイプ分け(企業、大学、それ以外の公的研究機関等の 3 分類)を行った。具体的には、PATSTAT をベースとした OECD-HAN (Harmonized Name) 情報を用いて、特許出願人の表記ゆれの修正と同時に出願人のタイプ分けを行う (PATSTAT の情報を USPTO データに特許単位で移送) (European Patent Office, 2016)。SCOPUS データについては著者所属先情報のテキストデータを用いてルールベースでタイプ分けを行った。なお、機関名称におけるタイプ別のキーワード (タイプ分けのルール) は、OECD-HAN 情報を元に抽出した。

次に論文著者の所属機関名と特許出願人の機関名 (なお、個人出願特許は除く) の接続をタイプごとに行った。両者のテキストマッチは Token Base の Approximate Match による。Token (単語) の頻出数の対数値の逆数でウェイトをつけて、両者の Token ベクターのジャカード指標を用いて、一定の閾値以上のものをマッチさせるようにした。なお、閾値の設定については、機関のタイプ別、機関名のワード数によって接続状況を見ながら調整した。また、氏名については Token Base のウェイトなしのジャカード指標を用いて Approximate Match の情報を作成した。最終的に機関名と氏名の両方がマッチしているレコードを同一人物であると判断した。

SCOPUS と USPTO データの著者・発明者マッチの結果を表 1 に示す。結果の読み取りの前に今回の作業に関する留意点をいくつか述べておきたい。まず、表 1 における著者 ID 数は実際の研究者の数を示しているものではない。SCOPUS のデータベースにおいては著者 ID が存在するが、同一人物に対して複数の著者 ID が振られている (Splitting Error が大きい) 傾向があることが分かっている。一方、USPTO データには、発明者識別作業を行った結果としての発明者 ID 情報が存在する。発明者の識別作業とは、複数の特許レコードに存在する同一発明者を識別する作業である。具体的には、著者の氏名その他、特許の権利人や住所情報、特許の技術分類や共同発明者などの情報を用いて、複数特許レコードの発明者が同一人物であるか否かの確率モデルを作成して、その結果をベースにクラスタリング作業を行うことになる。米国においては USPTO の特許データを用いた研究が行われてきており (Li et al., 2014)。現在では USPTO がこの作業を内生化し、特許情報をアップデートするたびに発明者の識別作業を行い、その結果を公表している。² 本研究においては、この USPTO の情報を用いているが、この情報はあくまで推計結果なので、表 1 における Inventor 数が実際の発明者数と一致しているかどうかは分からない。表 1 によると特許データとマッチで

² 日本のデータを用いた先行研究においては JPO 特許レコードの発明者識別作業を行った上で、SCOPUS 著者情報とのマッチを行っている (Ikeuchi et al., 2017)。

きた SCOPUS 著者数は 190,892 であるのに対して、論文データとマッチできた USPTO 発明者数は 168,472 となっている。つまり、SCOPUS の方が同一発明者に対して複数の ID が紐づいている Splitting Error が大きい傾向にある。

(表 1)

また、SCOPUS の著者と USPTO の発明者のリンケージは、M 対 N の形式になっており、両者の最大公倍数をとって研究者単位の情報にはしていない。その理由としては、ここでの分析の目的が、サイエンスとイノベーションの共起化現象のトレンドを明らかにすることにあり、個々の研究者単位で分析を行うことにはないからである。表 1 でみるように、SCOPUS の著者数全体において特許発明者でもある著者の割合は 4.7% という結果を得たが、この数字は Author_ID 数のバイアス (Splitting error の割合) が特許にも関与している著者とそうでない著者で同じであると考えると真の値に近いものとなっているはずである。なお、一人の発明者情報とマッチできた複数の Author_ID を集約して一人とすることはできても、それに対応する発明者とマッチできない著者を集約する情報は存在しない。更に、本研究においては論文数や特許数でサイエンスとイノベーションの共起化の状況を見ることとするが、その場合は、一つの論文 (または特許) に両方の活動に関与している研究者が存在するか否かが重要なのであり、異なる論文や特許で同一の研究者が著者や発明者となっているか否かという情報は必要ない。

このような前提条件のもと表 1 をあらためて見ると、まず論文著者については、全体の 4.7% が発明者でもあり、所属機関のタイプ別にみると大学が 6.9%、公的研究機関が 0.7%、企業が 17.4% となっている。営利企業において発明者割合が高くなるのは当然といえるが、17.4% という数字 (82.6% の企業所属の論文著者が特許を書いていない) は低すぎると考えられる。一つの原因としては、論文データと特許データがカバーしている期間にズレがみられることにある。例えば、出願年でみると特許データは 2011 年までしかカバーされないが、論文データは 2015 年までの公表論文が収録されている。従って、2011 年以降に研究活動に入って論文を執筆している研究者は特許レコードに収録される可能性がない。また、所属機関と氏名によるマッチングがうまくいっていないこともあるだろう。この点については、Approximate Matching の閾値を下げることでより多くのマッチングレコードを得ることができるが、その場合異なる人物が誤ってマッチされる確率 (False Positive) が増えることに留意が必要である。

次に特許発明者の状況について見ると、論文の著者でもある研究者の割合は、全体で 13.3%、所属機関のタイプ別では大学で 22.6%、公的研究機関で 9.2%、企業で 10.3% となった。研究成果として論文の公表が本務といえる大学の研究者において 22.6% という数字はやはり低すぎる。特許データの場合、全体の 300 万件のうち 100 万件が 1996 年以前 (SCOPUS データの対象外) となっており、データ期間のズレの影響が特に大きくなっていると思われる。

る。

3. 接続データからみるサイエンスリンケージのトレンド

図2は出版年別に見た論文数について、特許発明者である著者を含む論文とそれ以外の論文について見たものである。前者の論文はイノベーション（特許）との共起化が見られる科学的知見のアウトプットとみることができるが、この論文数は2005年まで上昇したあと伸び率が低下している。一方で、特許発明者を著者として含まない論文数は高い伸び率が続いており、その結果として特許発明者を著者とする論文の割合（折れ線グラフ）は低下傾向にある。

（図2）

図3は出願年別に見た特許数について、論文著者を発明者として含む特許とそれ以外の特許について見たものである。こちらについては、前者の上昇率が後者を上回っており、論文著者を発明者として含む特許（サイエンスとの共起化が見られる特許）の割合が高まっている。論文数の増加によって科学的知見の拡大が進む中で、イノベーションとの共起化が見られる科学的知見の割合は低下する一方で、イノベーション（特許）からみたサイエンスリンケージに割合は高まっている結果となった。なお、日本の研究者に対する先行研究では、2000年以降、サイエンスとイノベーションの双方からみたリンケージ割合が高まっているという結果となっており、サイエンスサイドからみた傾向が米国と異なっている（Ikeuchi et al., 2017）。日本においては、論文アウトプットのマジョリティを占める国立大学の法人化が2004年に行われ、大学からの特許出願が急増した（Motohashi and Muramatsu, 2012）。論文数に占める特許発明者の割合の上昇は、この制度的要因の影響が大きいと考えられる。一方で、特許からみたサイエンスリンケージは日米両国とも高まっている。つまり、両国においてイノベーションにおける科学的知見の重要性は高まっている。

（図3）

最後に著者・発明者接続データからみたサイエンスリンケージの状況を、従来の手法である特許の特許文献引用から作成される指標と比較した結果を示す。なお、特許で引用されている非特許文献は、科学論文だけでなく特許のアブストラクトや一般的な書籍も含まれている。特に特許のアブストラクトについては特許そのものを引用していることと内容的な違いがないので、これらの文献を取り除くことが重要である（Callaert et al., 2006）。ここではUSPTOのデータベースにおける特許から引用されている非特許文献から、テキストマッチによって特許のアブストラクトを取り除いた（約3000万の非特許文献のうち約700万件を削除）。特許単位で論文著者が発明者として含まれているものと（特許アブストラク

トを除いた) 非特許文献の引用を行っているものの割合を出願年別、特許の技術分野別に比較したものを図3、図4に示す。

(図3)、(図4)

まず時系列でみた両者のトレンドであるが、非常によく似たパターンとなっている。なお、非特許文献引用による指標でみると最近のサイエンスリンケージの上昇傾向はより明確である。技術分野別の状況についても概ね同様の傾向を示している。“Micro Structural and nano technology”, “Pharmaceutical”, “Biotechnology”, “Organic fine chemistry”などのサイエンス型技術といわれる分野で両者の割合が高くなっている。ただし、“IT methods for management”のように非特許文献割合が高い一方で、研究者に着目した指標についてはそう大きくない技術分野も見受けられる。この傾向は“Computer Technology”や“Digital Communication”などの情報通信分野に見られる。これは非特許文献にソフトウェアに関するドキュメントが含まれていることの影響であると推測される。非特許文献には多様な文書が含まれているのに対して、論文著者を発明者とする指標は科学論文のみを対象としたものとなっている。前者はサイエンス→イノベーションというリニアモデルを前提としたもの、後者は両者の共起化を示したものと内容的に違いがあるとともに、対象としている文献の違いにも図3及び図4の両者の乖離の原因となっている。

4. AI分野における詳細分析

ここではAI分野にフォーカスして、サイエンスとイノベーションの共起化についてより詳細に分析することとする。AI分野は、近年、ディープラーニングなどの機械学習の手法の発展が著しく、その一方で様々な産業分野でのイノベーションとしての実装も進んでいる。イノベーションの基礎となる手法の開発はコンピュータサイエンス分野のアカデミアを中心として進んでいるが、その実装は産業界で行われる。また、グーグルやマイクロソフトなど企業における研究者が研究論文を公開するという見受けられ、サイエンスとイノベーションの共起化を見るうえで適当な事例と考えられる。また、ディープラーニングを使った医薬品の開発など、AIが新しいイノベーションを起こすための手法(IMI: Invention for Method of Invention)ともいわれており(Cockburn et al., 2018)、コンピュータサイエンスの一分野であるAIが他の技術分野も含めたイノベーション全体への与える影響についても見ることにする。

まず、論文と特許についてAI分野を定義する。論文についてはSCOPUSデータベースにおいてElsevier社が提供するASJC (All Science Journal Classification)を用いる。ジャーナルごとに学術分野の分類を付与したもので、Computer Scienceの一分野としてArtificial IntelligenceがASJC=1702と定義されているので論文についてはこの分類を用いる。特許については、IPCコードのG06Nを用いた。この分類は特許庁の平成26年度技術動向調査「人

工知能分野」で用いられているものであり、「ニューラルネットワーク、推論マシン、フuzzyロジック等の特定の計算モデルを活用したコンピュータシステム」と定義されている（特許庁、2015）。具体的には、ディープラーニングを含むニューラルネットワークや推論システムなどの基本的な分析モデルや手法、ヒューマンインターフェース（データ可視化など）、画像やテキストなどの認識、データマイニング、検索などの情報処理、予測や機器のコントロールなどの技術が含まれている。なお、OECDのSTIスコアボードにおいては特許データを用いたAI技術に関する統計が掲載されているが、OECD（2013）に基づく一般的なソフトウェア（G06F）を含む幅広い分類といる。それとの比較では、本論文におけるAI分野や基盤的技術に限定した狭義の分類とすることができる。

図5はAI論文とAI特許の全体に占める割合をしめしたものである。論文と特許の両者とも2010年までは上昇傾向にあるが、その後伸びはとまっている。なお、全体に占めるAI論文の割合は1%以下、AI特許については0.2~0.3%となっている。また、特許については、USPTOの情報が登録特許のみとなっていることから、出願年で見ると2011年までの状況しか窺い知ることができない。

（図5）

次に論文の著者についてその所属機関のタイプ別に見た。図6は企業に所属する著者の論文シェアを見たものである。全体的なトレンドとしては下降傾向にあり、企業の基礎研究からの撤退を裏付ける内容となっている（Arora et al., 2015）。すべての学術領域について、特許発明者を著者とする論文については、企業論文の割合が15%程度と全体（10%以下）と比べて高くなっている。特許活動は企業セクターが中心となるので、この結果は当然といえるが、この割合についても最近減少傾向にある（逆に、企業以外、つまり大学や公的研究機関の特許発明者を含む論文シェアの高まりを示している）。AI論文について見ると、企業論文シェアは論文全体とほぼ同等の数字となっており、やはり傾向的には企業論文は減少傾向である。しかし、特許出願人を著者とするAI論文については、企業割合は20%前後と比較的高い水準で推移しており、特許に関連する（イノベーションに近い）学術論文については、企業研究者も積極的に取り組んでいる様子がうかがえる。

（図6）

同じ趣旨で特許サイドから見たものが図7である。ここでは企業以外の大学・公的研究機関の特許割合を見ている。まず全体技術分野で見るとほぼ横ばいで推移しているが、論文著者による大学等の特許割合は減少傾向にある。企業としては基礎研究から手を引きつつあるが、特許とからむ論文執筆については手を引いているわけではない。この傾向はAI特許についてより明確である。AI特許全体、論文著者を発明者にもつAI特許の両者において、

大学等のシェアの減少、つまり企業特許シェアの増加が見られる。

(図7)

表2はこれらの結果をまとめたものである。ここではAI論文、特許の全体に占める割合を、著者(発明者)の所属機関の2タイプ×著者・発明者によるものか否かの4つの分類で見たものである。すべての研究者は、①論文の著者のみ、②論文の著者であると同時に特許発明者、③特許の発明者のみ、の3種類に分類することができるが、まず企業の研究者について見ると、AI論文の割合は②論文・特許の場合0.91%、①論文のみの場合0.55%で、②>①となる。また特許について見ると③特許のみの場合0.15%、②論文・特許の場合0.11%となり、③>②となる。つまり、企業においては、③>②>①となり、AI分野については特許のみの活動が最も活発に行われているということである。同様に大学等に見ると、論文の割合は②>①と企業と同様であるが、特許については③<②となる。つまり、大学等においては論文・特許の両方に従事する研究者割合が相対的に高いことを示している。

(表2)

最後にAI分野の他の技術分野への広がりについて見た。図8は、著者・発明者の両方の研究者において、AI発明者(AI特許を発明したことがある研究者)とAI著者(AI論文を公表したことがある研究者)が、どのような技術分野で特許を発明しているか示したものである。なお、ここではAI特許が含まれるコンピュータサイエンスの分野を除いた技術分野について全体に占める特許数シェアを示している³。まず、両者とも幅広い技術分野に特許が広がっているとともに、この傾向はAI著者の方が強いことが分かる。科学論文の方がより基礎的な内容で応用範囲が広いことによるものと考えられる。AI発明者と比べてAI著者においてシェアが高くなっている技術分野はMedical TechnologyやOrganic Fine Chemistryであるが、このようなライフサイエンスの分野においてもAIサイエンスをベースとする特許が存在しており、AIがIMI(Invention of method of invention)であることを示している。

(図8)

³ なお、コンピュータサイエンスの特許シェアは、AI発明者について53.2%、AI著者について41.6%である。

5. まとめと今後の研究課題

本研究においては、米国の企業・研究機関に所属する研究者について、科学技術論文と特許データを著者・発明者レベルで接続を行い、研究者単位でみたサイエンスリンケージのトレンドについて調べた。また、AI分野にフォーカスして、サイエンスとイノベーションの共起化現象に関して、より詳細な分析を行った。

まず、米国においては論文数でみたサイエンス（科学的な新しい知見）の広がりが見られる。そのうち、特許発明者による論文数について、総数としては増加しているが全体の論文数に占める割合は低下している。これは特許活動を活発に行う企業が基礎研究から撤退する傾向にあり、企業著者による論文数が減少していることとも関係していると思われる。一方で特許からみたサイエンスリンケージ（論文著者による特許発明の割合）は増加傾向にある。イノベーションにおける科学的知見の重要性の高まりを反映した結果となった。

次にAI分野にフォーカスして詳細な分析を行ったところ、やはり企業著者による論文数シェアは低下傾向にあるが、特許発明者による論文数については企業シェアの低下は見られていない。また、特許からみた傾向についても、AI特許について、論文著者によるものであるか否かにかかわらず、企業発明者によるものの割合が増えていることが分かった。企業セクターにおいては、オープンに公表される科学技術論文にも取り組みながら、同時に特許による技術の囲い込みを行っており、その傾向が高まっていることが分かった。一方で、大学等の公的研究セクターにおいても、論文著者が特許活動にも乗り出す傾向にあり、オープンなサイエンスと特定の所有者による技術の権利化が同時進行で進んでいる姿が浮き上がった。

最後に今後の研究課題についてまとめることで論文を締めくくることとする。データベースについては様々な改善点が考えられる。まず、特許と論文の分析対象期間がずれていることから、アンマッチとなる研究者の割合が大きくなった。両者をそろえることで研究者に着目したサイエンスリンケージの状況をより正確にとらえることが可能となる。また、論文数や特許数については、それぞれの質を勘案することも可能である。特に論文数については学術誌数の増加に伴って増加傾向にある。従って、SCIMAGOやSCIE（Web of Science）などのIndexを活用して、一定水準以上にある学術誌の論文に限定して分析を進めることも検討すべきである。研究者に着目した新しいサイエンスリンケージ指標の評価をより正確に行うためには、特許が引用する非特許文献についても科学論文の抽出や個々の論文のQuality Indexの作成などの作業が必要となる。

また、分析の面では、サイエンスリンケージのダイナミクスについて検討が必要である。例えばAI分野については、企業セクターの特許による技術の囲い込みが進んでいることを示したが、学術論文によって公開される情報との関係はどのようになっているのか？よりオープンなサイエンスをベースにイノベーションが生まれているのか、あるいは、イノベーションから新しい科学的知見が生まれるという逆のループも存在するのか？その際に利益を求める企業と公的な研究資金で運営されている大学等との関係はどのようになっている

のか？これらの問題についてはこれまでの様々な角度から論じられてきているが (Van Looy et al., 2003; Magerman et al., 2015)、論文と特許を接続したデータベースは新しい切り口の分析を可能とするものとして期待できる。

参考文献

- 特許庁 (2015), 『平成 26 年度特許出願技術動向調査報告書 (概要) 人工知能技術』 特許庁.
- 元橋一之 (2014), 『日はまた高く 産業競争力の再生』、日本経済新聞出版社.
- Arora, A., S. Belenzon, and A. Pataconi (2015), “Killing the Golden Goose? The Decline of Science in Corporate R&D,” *NBER Working Paper* No.20902.
- Callaert, J., B. Van Looy, A. Verbeek, K. Debackere, and B. Thijs (2006), “Traces of Prior Art: An analysis of non-patent references found in patent documents,” *Scientometrics* 69 (1), 3-20.
- Cockburn, I., R. Henderson, and S. Stern (2018), “The impact of artificial intelligence on innovation,” *NBER working paper* No.24449.
- European Patent Office (2016), The Worldwide Patent Statistical Database (PATSTAT), Autumn edition.
- Ikeuchi, K., K. Motohashi, R. Tamura, and N. Tsukada (2017), “Measuring Science Intensity of Industry using Linked Dataset of Science, Technology and Industry,” *RIETI Discussion Papers Series* 17-E-056.
- Li, G. C., R. Lai, A. D’Amour, D. M. Doolin, Y. Sun, V. I. Torvik, A. Z. Yu, and L. Fleming (2014), “Disambiguation and co-authorship networks of the US patent inventor database (1975–2010),” *Research Policy* 43 (6), 941-955.
- Magerman, T., B. Van Looy, and K. Debackere (2015), “Does involvement in patenting jeopardize one’s academic footprint? An analysis of patent-paper pairs in biotechnology,” *Research Policy* 44 (9), 1702-1713.
- Motohashi, K. and S. Muramatsu (2012), “Examining the university industry collaboration policy in Japan: Patent analysis,” *Research Policy* 34 (2), 149-162.
- Narin, F., and E. Noma (1985), “Is technology becoming science?,” *Scientometrics* 7 (3)-(6), 368-381.
- OECD (2013), “Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by Big Data,” *OECD Digital Economy Papers* No. 222.
- Schmoch, U. (1997), “Indicators and relations between science and technology,” *Scientometrics* 38 (1), 103-116.
- Van Looy, B., V. E. Zimmermann, R. Veugelers, A. Verbeek, and K. Debackere (2003), “Do science-technology interactions pay off when developing technology?,” *Scientometrics* 57 (3), 355-367.

表 1 : 米国研究者に関する SCOPUS ・ USPTO 接続結果

	SCOPUS			USPTO			Matched
	Author_id	Matched	Share	Inventor_id	Matched	Share	Results
University	1,239,126	85,390	6.9%	310,861	70,338	22.6%	69,075
PRI	2,953,697	19,241	0.7%	187,413	17,308	9.2%	17,161
Firm	566,031	98,732	17.4%	900,476	93,043	10.3%	90,845
Total	4,065,766	190,892	4.7%	1,264,306	168,472	13.3%	163,997

表 2 : AI 特許と AI 論文の割合 (著者・発明者タイプ別)

	Share of AI patents		Share of AI papers	
	W_paper	WO_paper	W_patent	WO_patent
Private	0.11%	0.15%	0.91%	0.55%
Public	0.18%	0.08%	0.76%	0.60%

図1：特許発明者の有無と論文数の推移

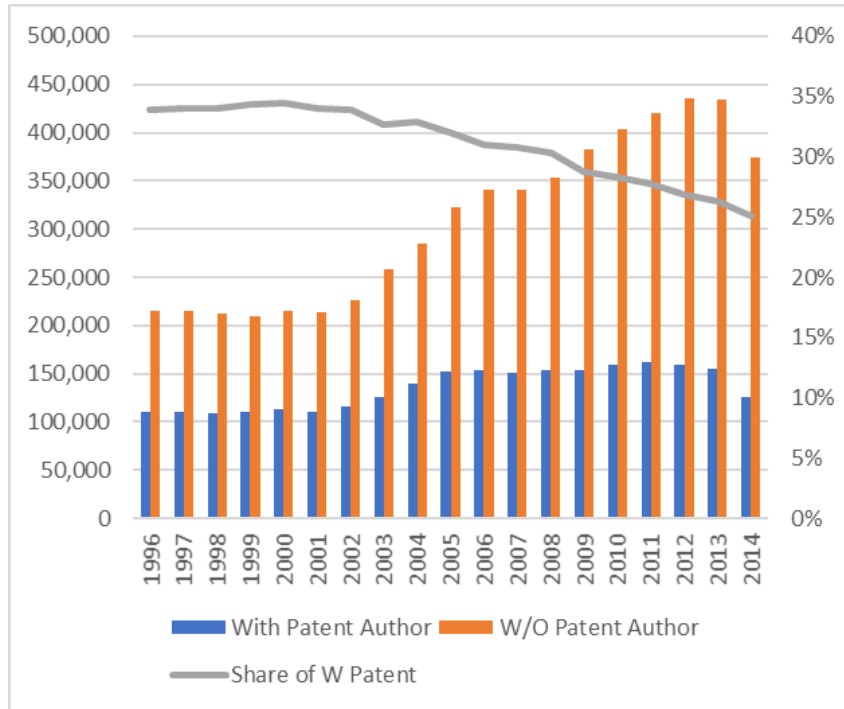


図2：論文著者の有無と特許数の推移

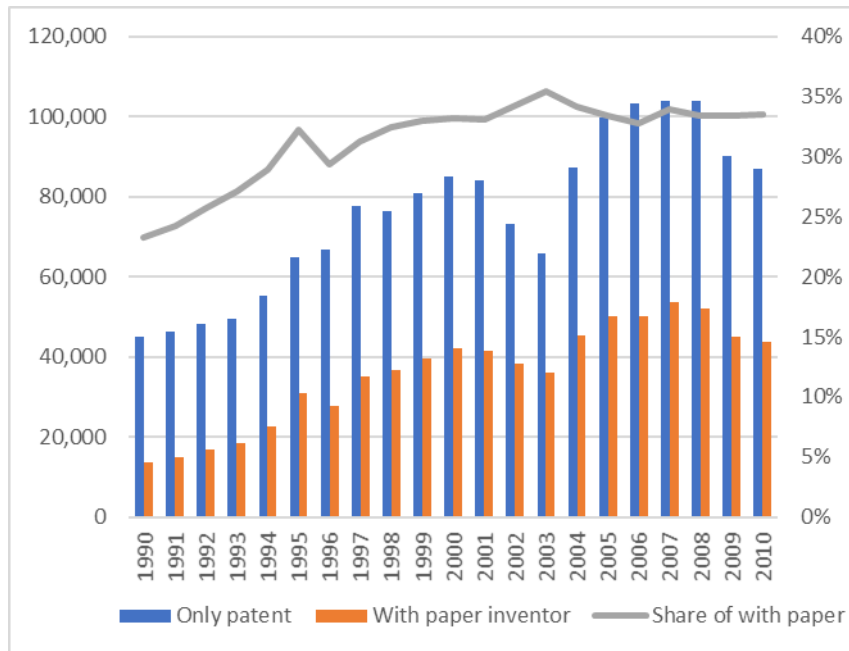


図3：非特許文献引用に関する情報との比較（時系列）

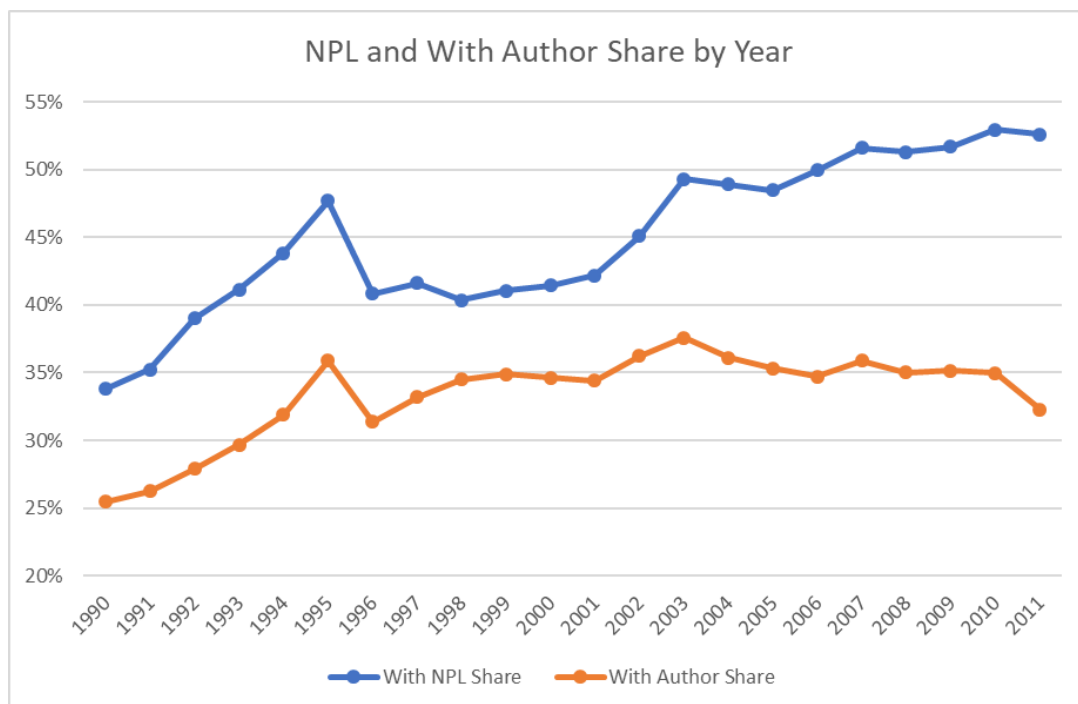


図4：非特許文献引用に関する情報との比較（技術分野別）

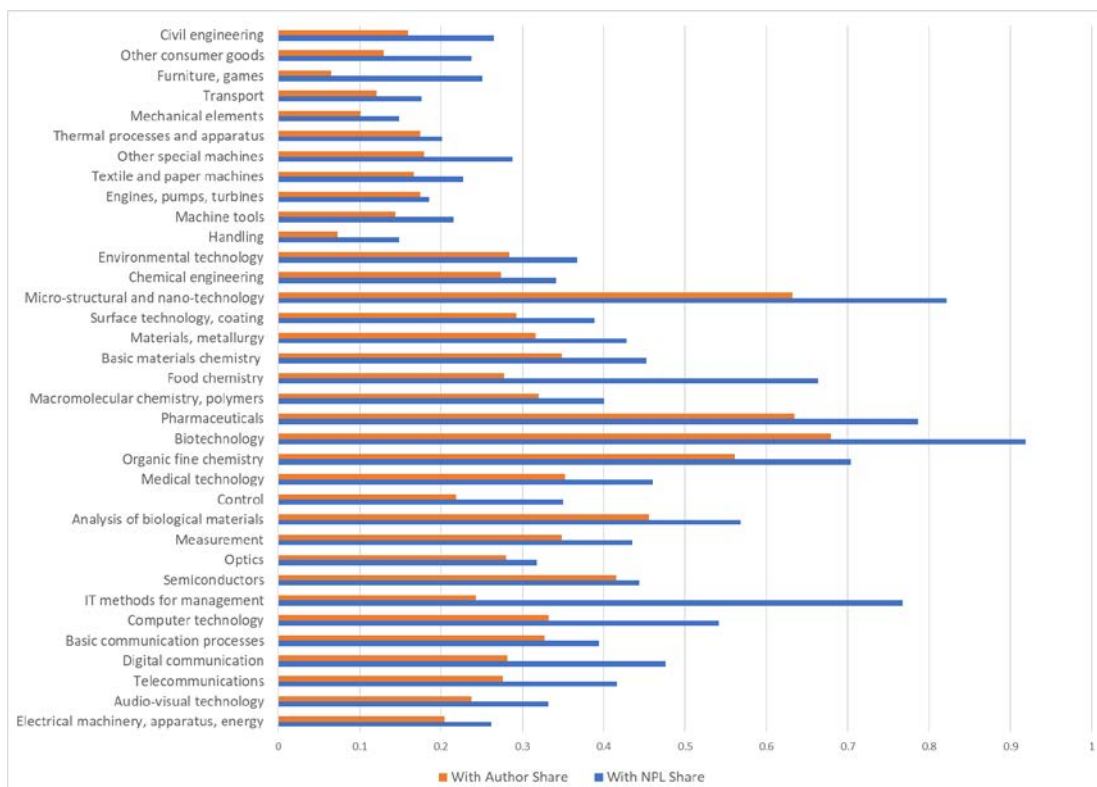


図5：AI論文、AI特許のシェアの推移

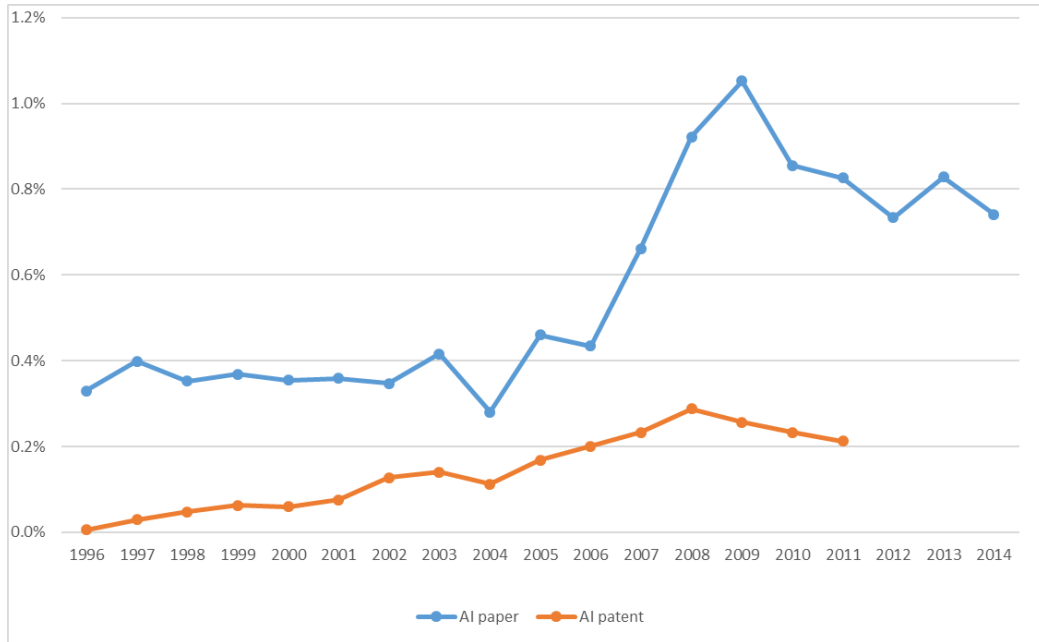


図6：企業研究者による論文シェア

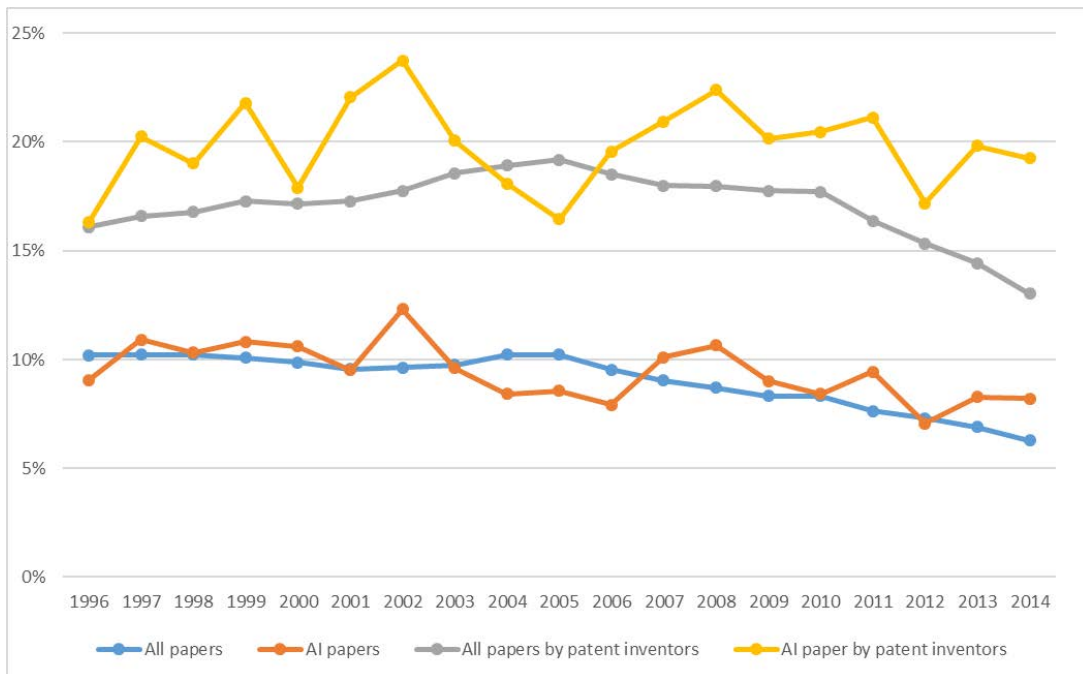


図7：公的研究機関の研究者による特許割合

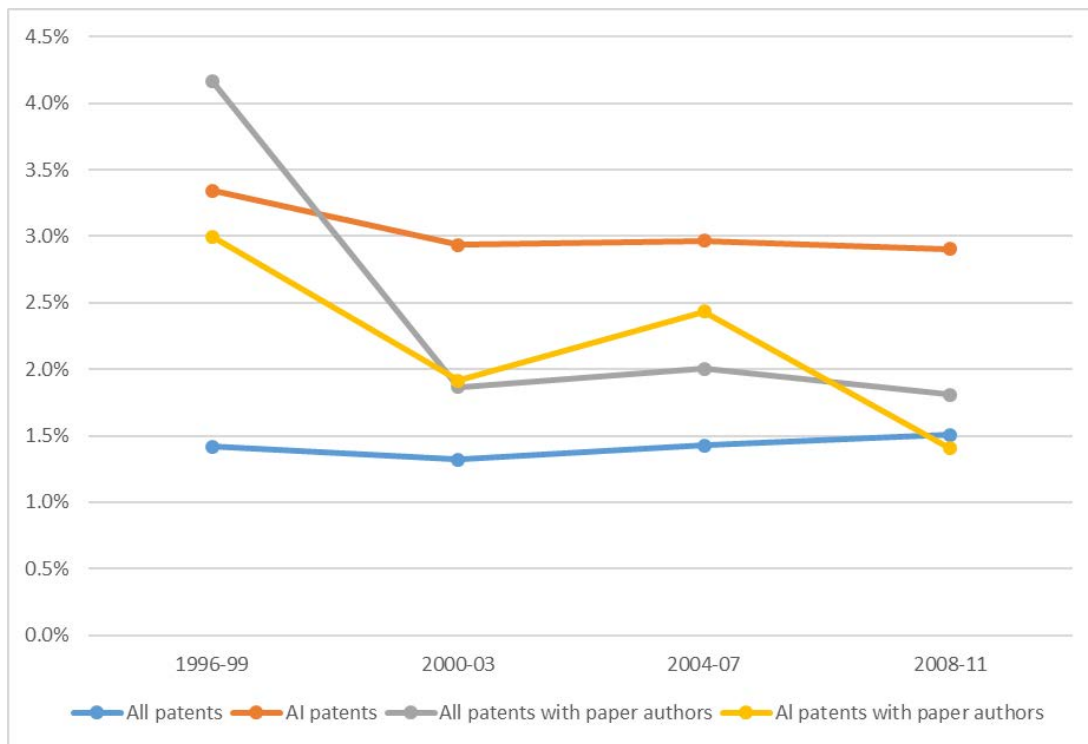
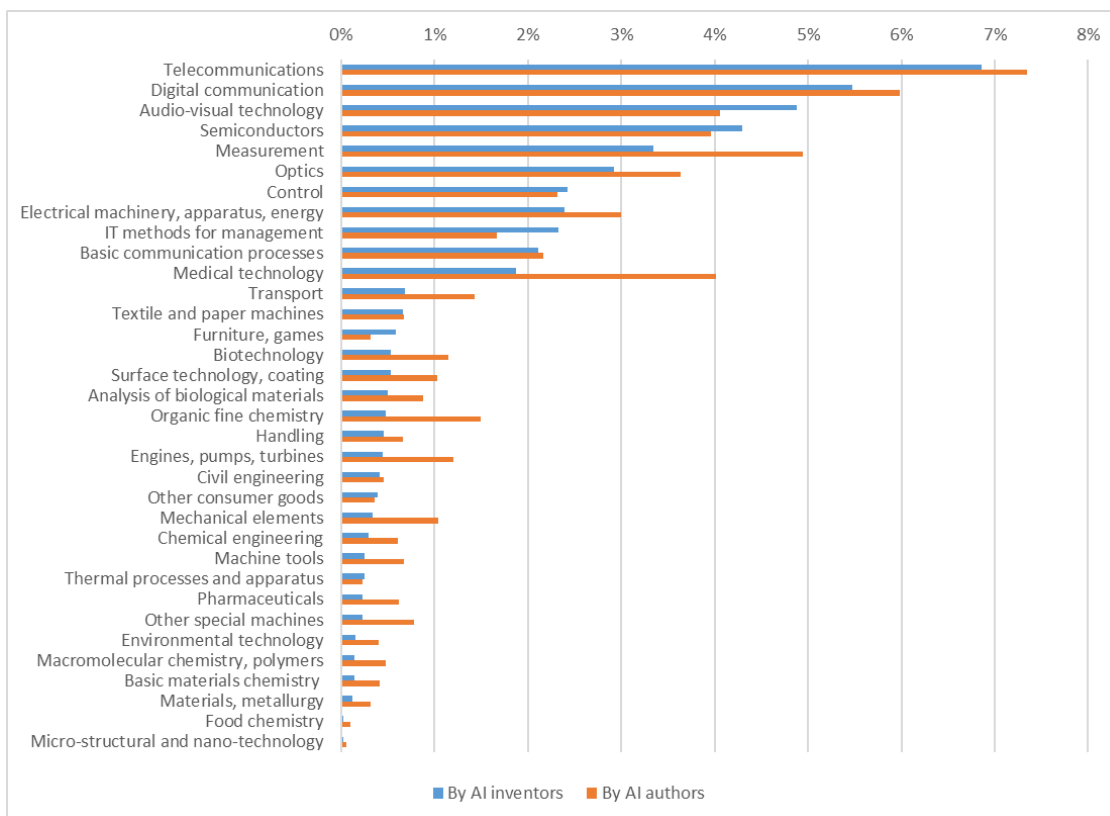


図8：AI 著者、AI 発明者による特許技術分類（コンピュータ技術を除く）



DISCUSSION PAPER No.160

AIにおけるサイエンスとイノベーションの共起化：
米国における論文・特許データベースを用いた分析

2018年7月

文部科学省 科学技術・学術政策研究所 第1研究グループ
元橋 一之

〒100-0013 東京都千代田区霞が関 3-2-2 中央合同庁舎第7号館 東館 16階
TEL: 03-3581-2396 FAX: 03-3503-3996

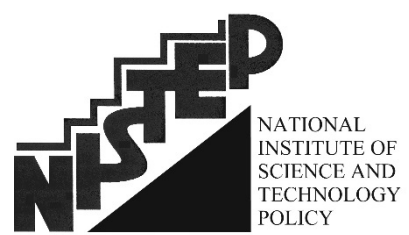
Co-occurrence of Science and Innovation in AI:
Empirical Analysis of Paper-patent Linked Dataset in the United States

July 2018

Kazuyuki Motohashi

First Theory-Oriented Research Group
National Institute of Science and Technology Policy (NISTEP)
Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan

<http://doi.org/10.15108/dp160>



<http://www.nistep.go.jp>