

Title	発明の同期の定量化の試み
Author(s)	田中, 秀穂; 金子, 永基; 渡辺, 亮一
Citation	年次学術大会講演要旨集, 35: 785-788
Issue Date	2020-10-31
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/17448
Rights	本著作物は研究・イノベーション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 Research Policy and Innovation Management.
Description	一般講演要旨

発明の同期の定量化の試み

○田中秀穂、金子永基、渡辺亮一（芝浦工業大学）
ssry16@shibaura-it.ac.jp

1. はじめに

同時期に独立した複数の人物により同じ発明・発見がなされる現象は、同期発明・発見、多重発明・発見、同時発明・発見などと称されてきた。本稿では「発明・発見の同期」という表現を用いる。発明・発見の同期は、イノベーション研究や技術予測や技術の進化論的解釈を進める上で重要な指標となる可能性がある。一方で、この発明・発見の同期の定義・基準は曖昧で、網羅的、定量的、客観的に同期の発生を計量することはこれまで十分になされていない。本発表では、発明の同期の定量化の試みについて報告する。

2. 発明・発見の同期現象

同時期に独立して複数の人物によって同じ発明、発見がなされる現象は、iPS 細胞（田中 2008）や変形性関節症治療法（Lubowitz ら 2018）などの最近の先端研究分野で観察されている事例のみならず、古くから多く発生してきたことが知られている。Ogburn and Thomas (1922)は、同時発明または発見の大規模なリストを作成し、148 件の同時期に複数の発明者または発見者が見いだされる案件を示した。この中には、古くは 17 世紀における太陽の黒点の発見や微積分の発見などから、第二次産業革命における様々な発明や 1900 年の遺伝法則の再発見などが含まれている。その後も Merton (1961)による 264 件のリスト、Simonton (1979)による 579 件の発明・発見の同期事例のリストが公表されてきた。このように発明、発見の同期が多数確認されていることから、これらが孤高の天才によってもたらされる偶発的な事象ではなく、隣接可能性領域の中で進化的な必然によって起こるものであることを示唆するものとされている。しかしこのようなリストの作成は、類似性や重複発見の独立性をどのように定義するのかという課題における合意が難しいことから、困難で、かつその客観性には常に議論が付きまとってきた。

また Bikard (2020)は、PubMed に 1960 年から 2018 年の間に登録されている約 2900 万件の論文データを用いて、キーワードの一致と共引用を指標として、同時発見である確率の高い論文対を抽出した。しかし、この手法には依然として、PubMed に掲載されているバイオ系の発見の分析に限定されるという課題が残る。さらには、引用を指標とするため、時期が一定程度以上の過去の同期ケースの検出に限られる。

一方で特許文献への記載をもとに、発明の類似性もしくは同一性を判定することも、特に産業界などにおいて行われてきた。企業にとって、価値ある同一の発明を他社に先駆けて出願することは重要な命題であり、自社の発明と同一の発明が他者から出願されていないか、多くの企業で継続して調査が行われている。そのような調査の結果が公表されることは少ないが、宗定 (2011) は、物質単位で発明が特定しやすい化学発明における事例を公表している。これによれば例えば、難燃性繊維の発明において、同一の発明が 1 年 5 か月の間に独立した 5 社から出願されていた事例や、熱可塑性プラスチックの発明において、同一の発明内容が 1 日違いで独立した 2 社から出願されていた事例などが示されている。この事例は、激しく厳しい企業間での技術競合の状況を垣間見せている。

また、Kingston (2004) は、米国のインターフェアレンス手続きに着目して同時発明の体系的分析を試みた。毎年約 200 件のインターフェアレンスが発生する中から、270 件の事例を調べ、アカデミックな学会における先端情報をもとにした大企業間、それも国際間での技術競合の激しさを明らかにし、特に日本企業の発明力の高さに言及している。インターフェアレンスはまさに同一の発明の有無を裁判手続きにより詳細に検討するもので明確に発明の同期を検出するものであるが、この制度は、2013 年

の改正特許法の施行による先願主義への移行によって、近年の分析はできなくなった。また裁判手続きが開始された事例のみが分析の対象となるという限定もある。

近年では、特許文献中の単語群をベクトル化して、複数の特許の類似度を計算する方法が知られている。この手法では特許文献の類似度に基づくクラスタリングなどが可能となる一方、発明の同期を検出する手法としては、十分ではない。また人工知能を用いた類似度の評価が検討されているが、類似度判定の基準が明確にならない可能性がある。

このように特許データをもとにした発明の同期に着目することで、比較的高い精度で同期を評価することができる可能性がある。一方で、網羅性、継続性という点では、これまで技術分野や時期が限定されたものしか事例の収集ができておらず、基準の明確性という点でも課題が残っている。

3. 発明の同期の指標化の取り組み

そこで本研究では、これらの課題を解決するために、特許文献中に現れる初出単語に着目する。科学、技術用語は、新しい概念が生みだされるとともに創作され、また特定の技術分野で生み出された技術用語が他の分野で発明に関連する用語として用いられる。この初出単語を検出し、その後一定期間の間に再出した特許文献を、初出単語が現れた特許文献と同期した発明とみなす手法を検討する。特許文献に初出する単語が一定期間内に再出した場合をとらえて、その発生頻度を発明の同期の程度の代理指標とする試みである。

3.1. 技術用語の抽出と同期を検出するプログラム

特許文献は科学論文に比し、フォーマットが決まっており、共通して使用される単語表現は文献間で類似度が高いと考えられる。特に「特許請求の範囲」は発明の範囲が明確に記載される部分であるとともに、修辭的な表現が使用されることが少ないことから、多くが発明に本質的にかかわる単語から構成されている。

この「特許請求の範囲」から技術用語を抽出する方法としては、辞書を用いて分かち書きを実施し、固有の単語を得る方法が知られている。しかし発明の同期を検出するという目的においては、初出単語に着目することが必要であるため、辞書機能を使用することは適切でない。例えば、複合語形態素解析プログラム **KHCoder** を用いた方法では、十分に特許中の初出単語を抽出することは困難であった。実際に、IT分野の特許の特許請求の範囲から **KHCoder** を用いて用語を抽出すると、IPA シラバス中に記述されている IT 分野の技術用語を十分には検出できないことが確認されている。

そこで辞書機能を使用せず、「特許請求の範囲」のテキストを平仮名や句読点、記号を区切りとした単語を抽出することで、発明に関わる重要なキーワードを簡易に得る方法を検討し採用した。上記のアルゴリズムを組み込んだ **Java** 言語によるプログラム「同期検出プログラム」を作成した。また上記アルゴリズムで抽出されてしまう、様々な非技術用語を削除するために 2 文字以下の単語や、演算子が入る計算式を除去した。さらにこのアルゴリズムによって抽出された単語を、その出現頻度の降順に並べ、各単語の抽出元である特許文献の出願日、出願番号、出願人の情報を出力しできるものとした。このプログラムによる抽出結果の一部例を表 1 に示す。これに用いた特許データは、日立システムズが提供する **SRPARTNER** 国内・国外版によって、出願日が 1993 年から 2017 年の範囲の **IPC** サブクラスが **G06N** のデータをダウンロードして分析に供した。

表 1 には、得られた単語群の中で、特定の単語が記された最先の特許文献の出願日から一定期間（この例においては 18 か月間）以内に出願された、同じ単語が記された特許文献を収集した。これらの情報をもとに、発明に関わる初出単語が最先から一定期間内に出現した場合に、同期した単語と判定する。上記の例で言えば、「量子ビット」は 1998 年 1 月 28 日に日立製作所が出願した特許に初出し、およそ 1 年 2 か月後に日本電気により出願された特許に再出しており、18 か月間の基準を適用するならば、この二つの特許で同期したということになる。また 2 番目の「人工ニューロン」は、わずか 5 日間の間に相次いで出現しており、短い期間で同期したことが分かる。なお今回用いた 18 か月という期間は任意であり、より適切な期間は今後の検討によるが、特許出願公開期間内に出願された特許ということで、単なる模倣の可能性を排除するという観点からは一定の合理性があると考えられる。

上記により抽出された単語群が、発明の本質に関わる適切な単語であるかどうかを、**KHCoder** を用いて抽出した場合と比較しながら検討した。やや主観的な手法にはなるが、特許請求の範囲から目視で

発明の本質にかかわると考えられる単語を 408 語抽出し、これらが上記プログラムおよび KHCoder で抽出されるかを検討したところ、本プログラムは 370 語、KHCoder は 298 語抽出した。擬陽性の抽出割合は、本プログラムは 29.9%である一方、KHCoder は 55%であり、いずれも本プログラムが KHCoder より高い抽出精度を示した。

表 1 「同期検出プログラム」の出力結果例

単語名	特許文献情報例			
	出願日	1998/1/28	1999/3/25	
量子ビット	出願番号	特願平10-15367	特願平11-82019	
	出願人	株式会社日立製作所	日本電気株式会社	...
	出願日	1995/7/14	1995/7/18	
人工ニューロン	出願番号	特願平7-178403	特願平7-212330	
	出願人	株式会社日立製作所	クーパー・タイヤ・アンド・ラバー・カンパニー	...
	出願日	1995/7/14	1995/7/18	

この単語レベルでの同期が、実際の特許技術内容としても同期に該当するかどうかは今後の検討が必要である。しかし、単語レベルで同期しているとして抽出された特許の対の中にそのような事例は見つかっていない。例えば、「プロピオン酸メチル」という単語で同期しているとされた二つの特許内容を精査したところ、いずれもリチウムイオン電池の過充電に対処するための非水溶性溶媒についての発明であり、二つは「同一」の発明であると確実に判断できる文献であった。

3.2. 同期検出プログラムによる特定分野特許の分析例

次に、このプログラムをファセット分類の ZHV（ハイブリッド自動車）に分類される特許を用いて単語レベルの発明の同期がどのように検出されるか、同期検出プログラムを用いて分析を試みた。出願日は 1993 年から 2017 年の間の特許データを用いた。この間の ZHV 分類特許の総数は 23217 件であった。年ごとの総出願数を見ると、1996 年から 2008 年にかけて一貫して増加し、その後横ばいとなっていた。この特許データに対して、先の分析と同様に初出単語の出現から 18 ヶ月の間に、異なる出願人から同じ単語を含む出願がなされた場合をカウントした。結果を図 1 に示す。なお図は示さないが、この間の年ごとの全ての新語の出現数は、2008 年をピークに増加し、その後漸減した。これは総特許出願数の推移に相似している。

図 1. ZHV 分類特許における同期した技術用語数（単位：語）

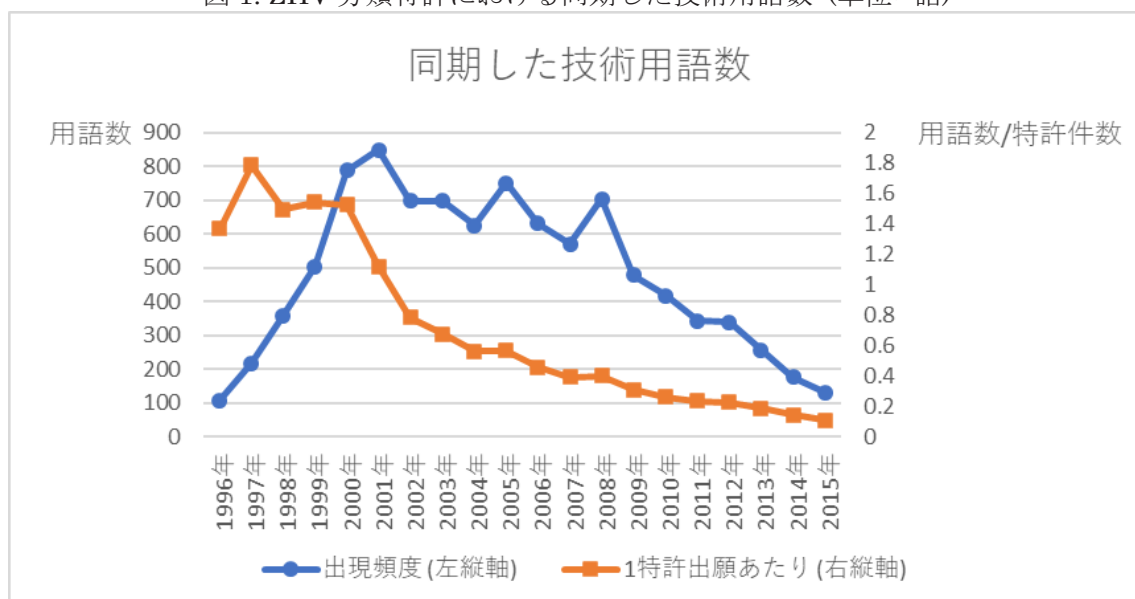


図1に見られるように同期した単語数は総特許数、全ての新語の出現数とは異なる推移を示した。まず同期した単語の総数は、2001年をピークとして増加し、その後漸減した。また1特許あたりの同期した単語数は2000年まで高い水準を示したのちに、その後一貫して低下した。いずれもそのピークが総特許数などよりも時期が早い。

これらの結果は、ハイブリッド技術分野の特許出願における技術競合は、比較的初期の段階で激しかったことを示唆する。トヨタの初代プリウスの発売は1997年であり、この前後の時期に初期の重要なハイブリッド自動車関連の基本特許が出願されたとされている。単語レベルでの発明の同期の程度は、総特許出願数がまだ少ないそのような時期に、基本発明の開発にトヨタ以外の企業も参入した状況を可視化する指標となっている可能性がある。

また1特許出願あたりの同期数が2000年までの初期に高く、値が1.0を超えていることは興味深い。この時期に発明の必然性が高まり、特許に新しいアイデア、技術が濃縮されていることを示唆する。またこの初期段階において同期、競合した、初出単語を含んだ出願を行った企業と、それに続いた企業の分析は、当該分野におけるその後の技術の優位性を予測する指標につながる可能性があると考えられる。

このように同期検出プログラムを用いた単語レベルでの発明の同期の検出は、長期間に渡る多様な技術分野での定量的、網羅的な分析に寄与することが可能であり、技術戦略、イノベーション研究などの様々な研究に寄与する可能性がある。

4. 今後の課題

単語レベルでの発明の同期という本指標の、イノベーション研究等における研究ツールとしての活用価値については、さらに検討が必要である。本指標が実際の発明の同期とどの程度の相関があるのか、十分に検討したうえで、技術の進化的側面(Ridley 2015, Johnson 2010)の分析においてその有用性を今後検討する。

方法論としての今後の検討課題としては、まずこの同期検出プログラムの精度を向上させる必要がある。本手法による単語抽出は、KHCoderに比しより発明の本質を示す単語が抽出できることは示されたが、抽出された単語の質をより客観的に評価し、さらに抽出精度を改善する必要がある。また、技術分野によって特許文献中の記述は大きく異なるので、この手法がどの分野で有効であるのか見極める必要がある。初出単語からどの程度の時間の再出を同期と捉えるかについても検討が必要である。また同期の判定において、本研究では独立した「出願人」単位で抽出を行った。技術競合を可視化するという面からはこれには一定の合理性があるが、一方で「発明者」単位での抽出の可能性についても検討の余地があると考えられる。

5. 参考文献

- [1] 田中秀穂 (2008) 「iPS細胞の科学, 技術, イノベーション」日本知財学会誌 Vol.5, No.1, 13-18
- [2] J. H. Lubowitz, et al, (2018) “Two of a Kind: Multiple Discovery AKA Simultaneous Invention is the Rule” The J. of Arthroscopic and Related Surgery, Vol 34, No 8, pp 2257-2258
- [3] W. F. Ogburn and D. Thomas (1922) “Are Inventions Inevitable? A Note on Social Evolution” Political Science Quarterly, Vol. 37, No. 1, pp. 83-98
- [4] Merton, Robert K. (1961). “Singletons and Multiples in Scientific Discovery: A Chapter in the Sociology of Science.” Proceedings of the American Philosophical Society 105 (5), 470-486.
- [5] Simonton, Dean K. (1979). “Multiple Discovery and Invention: Zeitgeist, Genius, or Chance?” Journal of Personality and Social Psychology 37 (9), 1603-1616
- [6] Michaël Bikard (2020), “Idea Twins: Simultaneous Discoveries as a Research Tool” Strategic Management Journal, Volume 41, Issue 8, Pages: 1528-1543
- [7] 宗定 勇 (2011) 「世界最高の特許庁を目指して」特技懇 no.263. 57-67
- [8] William Kingston (2004) Light on simultaneous invention from US Patent Office “Interference” records, World Patent Information 26 (2004) 209-220
- [9] Matt Ridley (2015), “The evolution of everything: How new ideas emerge”, Harper
- [10] Steven Johnson (2010), “Where good ideas come from: The natural history of innovation” Riverhead Books