

Title	発明の同期分析手法の検証
Author(s)	佐々木, 涼; 田中, 秀穂
Citation	年次学術大会講演要旨集, 37: 96-99
Issue Date	2022-10-29
Type	Conference Paper
Text version	publisher
URL	<a href="http://hdl.handle.net/10119/18682">http://hdl.handle.net/10119/18682</a>
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

## 発明の同期分析手法の検証

○佐々木涼, 田中秀穂 (芝浦工業大学)

## 1. はじめに

同一の発明・発見が、複数の人物によって、独自になされることが時折起こる。このような現象に対して田中ら[1]は「発明・発見の同期」という表現を用いた。発明・発見の同期を分析することで、イノベーション研究や技術戦略などの様々な研究に寄与する可能性がある。しかし、これまでに使用されている発明・発見の同期の定義・基準は曖昧であり、網羅的、定量的、客観的に計算する手法は未だ確立されていない。そこで田中ら[1]は特許文献中に現れる初出単語に着目する手法を考案した。特許文献中に初出する単語が一定期間内に再出した場合を捉えて、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。本研究ではこの同期分析手法を用いて医薬関連特許を対象に検証、改良を行い、分析を行った。

## 2. 先行研究

同時期に独立して複数の人物によって同一の発明、発見がなされる現象は最近の先端研究分野で確認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。Simonton(1979)は 579 件の発明・発見の同期事例のリストを公表し、発明の同期が発生してきたことを示した。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、新たな発明や発見は天才によってもたらされる偶発的な事象ではないことが示唆されている。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自性をどのように定義するかという問題があり、かつその客観性には常に議論が付きまどってきた。

また Bikard (2020)は、1960 年から 2018 年の間に世界の主要医学系雑誌に掲載された約 2924 万件の文献を用いて、キーワードベースの類似論文リストから、少なくとも 5 回以上共引用されているペアの論文を、同時発明である確率の高い論文のペアとしてデータを構築した。しかし、この手法は、世界の主要医学系雑誌に掲載された文献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定される。

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。そこで田中ら[1]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を検出し、その後 18 か月以内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考案した。特許文献中で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」は発明の範囲が明記される部分であると共に、修辭的表現が使用されることが少ないことから、多くが発明の本質に関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、テキストと平仮名や句読点、記号を区切りとした単語を抽出することで、発明に関わるキーワードを取得する方法が採用された。上記のアルゴリズムを Java 言語によって実装し、これを「同期検出プログラム」と名付けた。また、上記のアルゴリズムで抽出されてしまう、2 文字以下の単語や、演算子が入る計算式などの非技術用語を削除することで、プログラムの精度向上を試みた。更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できるものとした。

この同期検出プログラムによって検出された単語レベルの同期が、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検討の余地がある。しかし、同期しているとして抽出された特許の対の中にはそのような事例が存在することが示された。また、同期検出プログラムを用いた特定分野特許の分析として、ハイブリッド自動車技術に分類される特許から分析を行った。同期した技術用語数、1 特許あたりの同期した単語数の推移と、ハイブリッド分野の歴史を併せて分析を行うことにより、企業がハイブリッド技術の基本発明の開発に参入した状況を可視化する指標となる可能性を示した。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可能性があることを示した。

一方でこの同期検出プログラムの精度には向上の余地がある。適切な単語の抽出のための精度の改善、単語レベルでの同期が発明内容の同期とどの程度一致するのか、などを明らかにする必要がある。

### 3. 改善案の検討

本研究ではこの同期検出プログラムを用いて、医薬関連特許を対象に分析の検証を行った。医薬分野の特徴として、研究開発に多大な費用と長い研究期間を必要とされ、投資を回収しつつ利益を獲得するには特許による独占権を得ることが必須であることがあげられる。また、医薬関連の製品は他分野の製品に比べ、少数の特許から構成されている。そのため特許の重要性が非常に高い分野である。IPC(国際特許分類)のサブクラスに A61K(医薬品、歯科用又は化粧品用製剤の特許分類を示す)が付与されている特許を用いて単語レベルで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行った。出願日は 1988 年から 2019 年の間の特許データを用いた。

#### 3.1 発明分野の一致率の評価の検討

同期検出プログラムにより抽出された初出単語による同期が実際に技術内容としても同期した特許文献であるのかを確認するため、抽出された初出単語を含む特許文献とその 18 か月以内に再出した特許文献の筆頭 IPC が一致するか検証を行った。1993 年、1998 年、2013 年に出願された IPC に A61K が含まれる特許文献を無作為に 100 件抽出し、メイングループでの筆頭 IPC の一致率を算出した。

同期が検出された特許文献のメイングループでの筆頭 IPC の一致率は 25%前後と、1/4 程度しか一致しなかった。そこで、データセットを IPC に A61K が含まれる特許文献から、筆頭 IPC が A61K である特許文献に変更した。結果として、メイングループでの筆頭 IPC の一致率が、分析期間である 1998 年、2013 年において 50%弱となった(表 1)。

出願年	IPC に A61K が含まれる特許文献	筆頭 IPC が A61K である特許文献
1993 年	21%	40%
1998 年	27%	48%
2013 年	23%	47%

表 1.同期が検出された特許文献のメイングループでの筆頭 IPC 一致率

これにより、発明分野においても一定以上の一致率を確保することが可能となった。

#### 3.2 分析期間の検討

これまでのプログラムでは分析対象期間以前の 5 年間のデータ中に出現した単語を分析対象から除去することで、新語ではない単語を排除していた。1998 年からの分析対象とする場合、それ以前の直近 5 年間のデータを追加するために 1993 年からの特許文献をデータセットとすることとなる。本研究では 5 年間という期間が適切であるか検証を行った。1993 年から 2019 年の間に出願された特許のうち、IPC に A61K が含まれる特許文献を対象に、データセットの期間以前の直近 5 年間のデータである 1988 年から 1992 年に出願された特許文献を取得し、出現した単語のリストを作成した。作成した全 105543 単語のリストと、1993 年、1998 年、2013 年に出現した単語の比較を行った結果、1993 年では 57388 単語中 17828 単語(31.07%)、1998 年では 31652 単語中 2498 単語(7.89%)、2013 年では 12835 単語中 248 単語(1.93%)の初出ではない単語が含まれていた。1998 年と 2013 年の一致率には約 6%の差があり、分析期間内で抽出する初出単語の精度に差が生じてしまっていた。そこで、データセット分析対象期間以前の 10 年間のデータを追加して用いるよう変更を行った。

#### 3.3 名寄せ

これまでの分析では分析期間内で社名を変更した企業や、グループ会社同士の特許も異なる出願人とみなされ、同期した特許文献として検出されてしまうという問題があった。これは独立した複数の人物によって同一の発明がなされるという発明の同期の定義に反している。そこで、分析期間内に出願を行った出願人をリスト化し、名寄せを行うことで、同一企業内での同期データを排除した。

### 3.4 同期を反映しない単語の削除

これまでのプログラムで出力された新語には、発明の同期を反映しているとは言えない不適切な単語が検出されていた。そこでこれらの単語の接頭語、接尾語をリスト化し、リスト内の語を含む単語をインプットデータからすべて除去した。また、同期検出プログラムを用いて 2017 年に出版された筆頭 IPC のサブクラスが A61K の特許文献を分析し、検出された同期した特許の公開特許公報を実際に確認したところ、技術内容としても同期していた特許文献は 4 文字以上の単語で同期していることが判明した。そこで同期検出プログラムによって検出する単語の最低文字数を 3 文字から 4 文字に変更した。

## 4. 改善後の同期検出プログラムによる分析

上述のインプットデータの改善を行い、A61K 分野特許の同期検出プログラムによる分析を試みた。分析には日本国特許庁への出版文献を使用した。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・国外版によってダウンロードして分析に供した。まずは年ごとの特許出願件数と新語出現数の推移を分析した。結果を図 1 に示す。この間の筆頭 IPC が A61K である特許の総数は 54169 件であった。年ごとの特許出願数を見ると、2005 年をピークに 2010 年まで漸減、その後横ばいとなっていた。この特許データに対して、年ごとに出現した新語の数をカウントした。新語出現数は 2002 年をピークに 2010 年まで漸減、その後横ばいと、特許出願件数と新語出現数の推移が相似していることが示された。

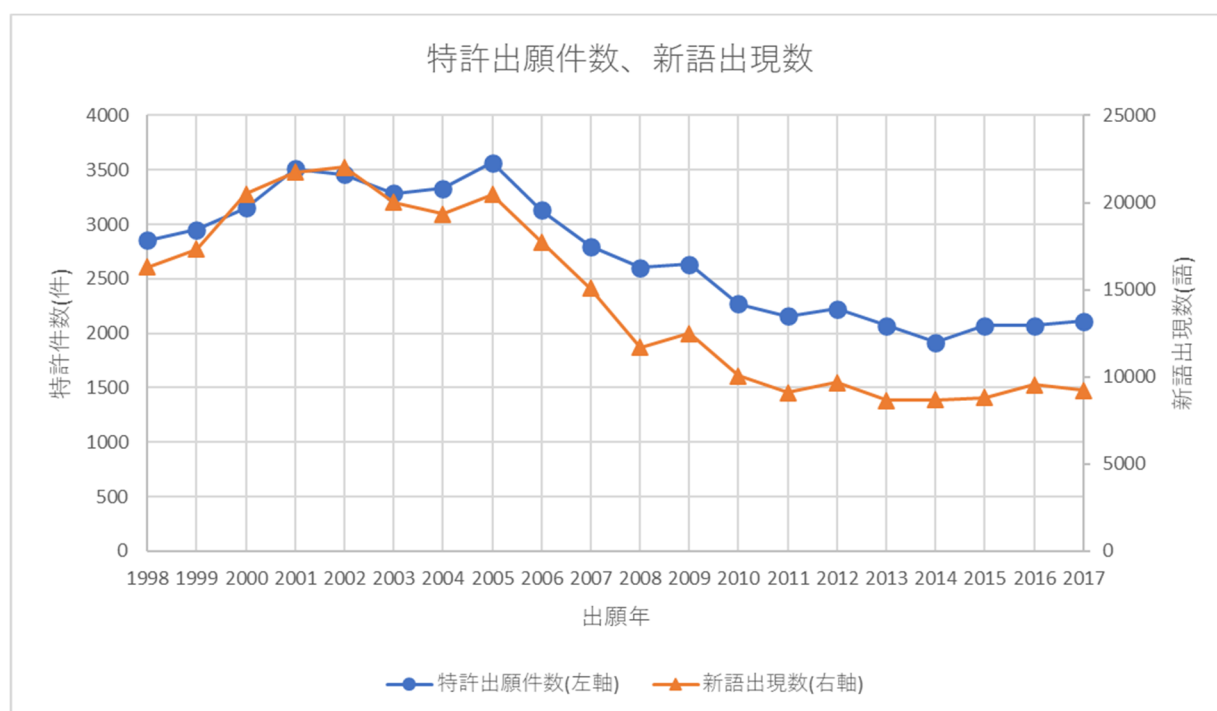


図 1 : A61K 分野における出願特許件数、新語出現数

図 1 より、特許出願件数が減少していることは、それだけ発明の権利化をする傾向が弱くなっていることを示唆する。また、新語出現数も特許出願件数と同じく減少している。特許文献中に現れる新語は新しい概念が生み出されるとともに創作されるため、特許文献における新しい概念の提示が減少していることを示唆する。

次に、A61K 分野における特許一件あたりの新語出現数と同期した単語数の推移を分析した。図 1 の分析と同様の特許データに対して、年ごとの新規出現数を特許出願数で割ることで特許一件あたりの新語出現数を算出した。また、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、同期した単語数を特許出願件数で割ることで、特許一件あたりの同期した単語数を算出した。結果を図 2 に示す。この間の特許一件あたりの新語出現数は、2000 年から 2002 年の間をピークに 2008 年まで漸減した後横ばいとなっていた。特許一件あたりの同期した単語数は 2000 年をピークに 2008 年まで減少、その後横ばいとなっていた。

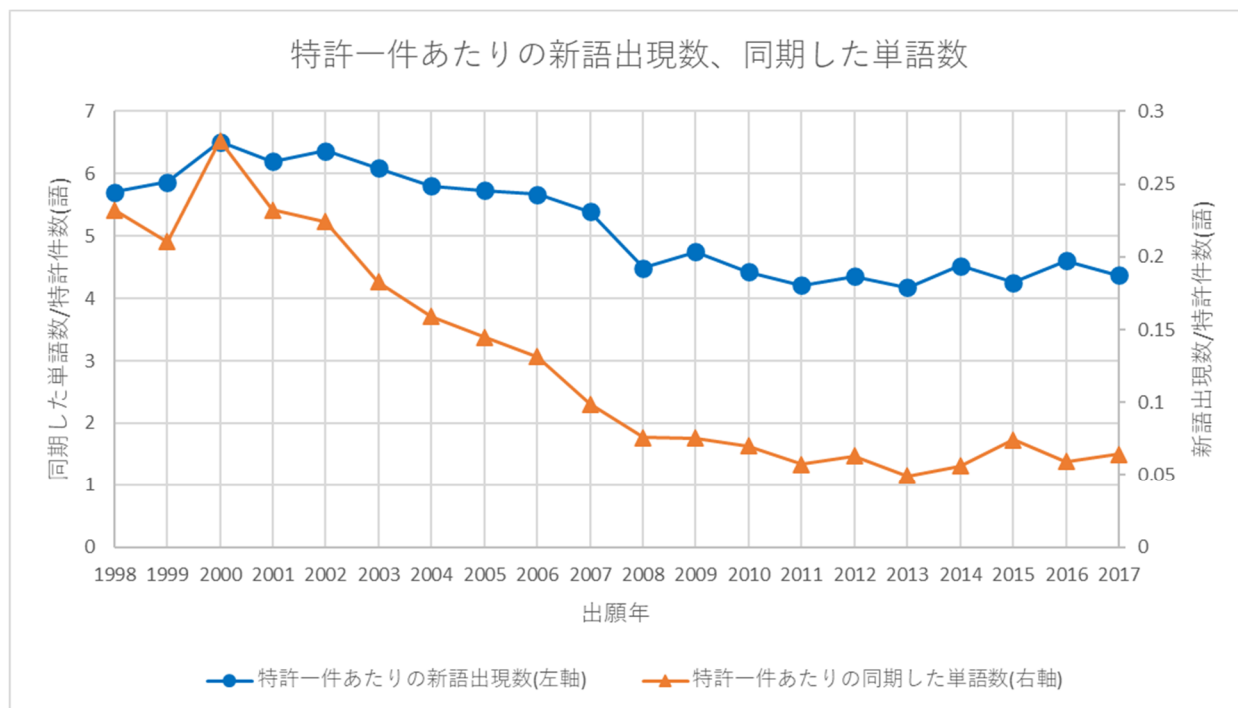


図 2 : A61K 分野における特許一件あたりの新語出現数、同期した単語数

図 2 より、特許一件あたりの新語出現数が減少していることは、特許一件あたりの新しい概念の提示数が減少していることを示唆する。しかし、特許一件あたりの新語出現数の値に着目すると、ピークである 2000 年から 2002 年の間は平均 6.4 語であるのに対し、2008 年から 2017 年の間は平均 4.4 語となっている。新語出現数の減少はピーク時の 7 割程度であり、変動は小さいと考えられる。また、特許一件あたりの同期した単語数が減少していることは、特許出願による技術競合は減少していることを示唆する。特許一件あたりの同期した単語数の値に着目すると、ピークである 2000 年は 0.28 語であるのに対し、2008 年から 2017 年の間は平均 0.06 語となっており、ピーク時の約 2 割まで値が減少している。よって、新しい概念の提示は微減に留まっているが、技術競合は減少しているということになる。このことから医薬関連分野において、近年研究テーマが分散している状況を読み取れると考えられる。

## 5. 今後の課題

単語レベルでの発明の同期という本指標が実際の発明の同期とどの程度相関があるのかについて十分に検討する必要がある。また、医薬関連分野以外の分野において、どのように単語レベルの同期が発生しているのか等を分析することで、本指標の活用価値について検討をする必要がある。また、本研究では「出願人」単位で発明の同期についての分析を行ったが、「発明者」単位での分析も検討の余地があると考えられる。

## 6. 参考文献

- [1] 田中秀穂、金子永基、渡辺亮一 (2020) 「発明の同期の定量化の試み」 研究・イノベーション学会 第 35 回年次学術大会 2G06
- [2] Simonton, Dean K. (1979). "Multiple Discovery and Invention: Zeitgeist, Genius, or Chance?" *Journal of Personality and Social Psychology* 37 (9), 1603–1616
- [3] Michaël Bikard (2020), "Idea Twins: Simultaneous Discoveries as a Research Tool" *Strategic Management Journal*, Volume 41, Issue 8, Pages: 1528-1543