

Title	論文著者IDによる研究者を単位とした計量書誌学的分析
Author(s)	川島, 浩誉; 富澤, 宏之
Citation	年次学術大会講演要旨集, 29: 373-375
Issue Date	2014-10-18
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/12466
Rights	本著作物は研究・技術計画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 Science Policy and Research Management.
Description	一般講演要旨

論文著者 ID による研究者を単位とした計量書誌学的分析

○川島浩誉, 富澤宏之 (科学技術・学術政策研究所)

本発表は計量書誌学に基づく研究者レベルの粒度を基本単位とする試行的分析を主題とする。近年、科学計量学においては分析単位のミクロ化が進んでいる。大学等の研究機関名のデータベース上の表記が整備・統一されるに伴い、伝統的な国レベルの集計・国際比較のみならず研究機関レベルの集計や研究ポートフォリオ分析が可能になった。さらに現在、各研究者に ID が付与されたデータベースによって研究者レベルの分析が可能になりつつある。本発表では、各論文の著者に ID が付与された論文データベースのひとつである Scopus を用い、日本の論文の停滞期における論文著者数の推移について報告する。

既に複数の先行研究で取り上げられているように、論文データベースを用いて集計された日本の論文生産数は 2005 年付近から横這い傾向が続いている (図 1)。集計の対象となる論文データベースに収録される論文誌数は年々拡大し続けていること、米国・英国・ドイツ等の主要国では日本が横這いを続けている期間も増加していることなどから、論文生産における日本の停滞として議論の対象になっている。

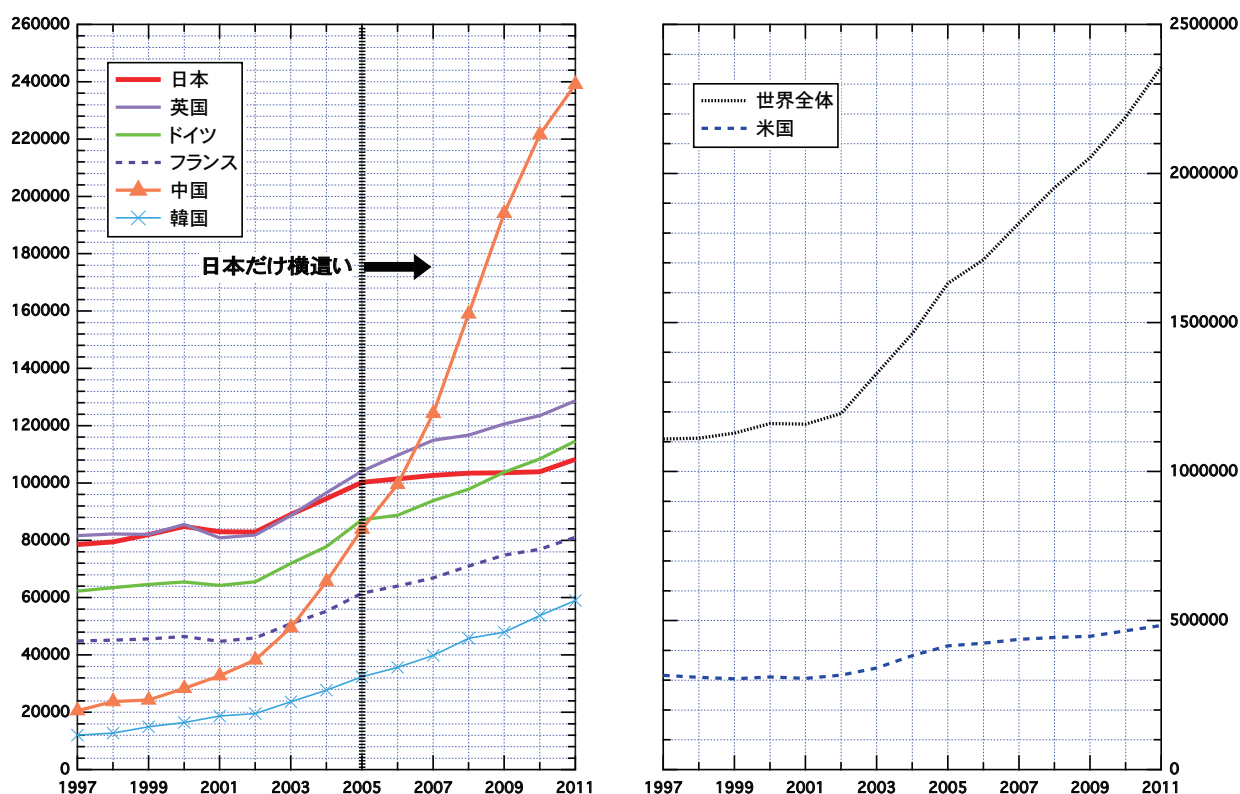


図 1 論文数推移 (英語論文のみを Scopus にて集計)

ここで、論文を生産する主体である研究者もしくは論文における著者という点からこの間の事象を考えると、1. 論文を書いて著者になる研究者が減っている 2. 一人あたりの論文生産数が下がっている、のいずれかもしくは両方である。本研究では 1 について検討する。

まず、論文著者の人数の推移を規模が近い英国とドイツと比較すると、図 2a、図 2b のように 2005-2007 年から 2008-2010 年に日本の著者数は微増に留まっている。著者のうち、各年に初めて論文に名を連ねたもののみを抽出したものが図 2c である。図 2c において、2005-2007 年から 2008-2010 年にかけては横這いであり、同期間の全ての著者数に関して表した図 2b と比べても増加率が小さく、論文数が横這いであることに関与していることを示唆している。

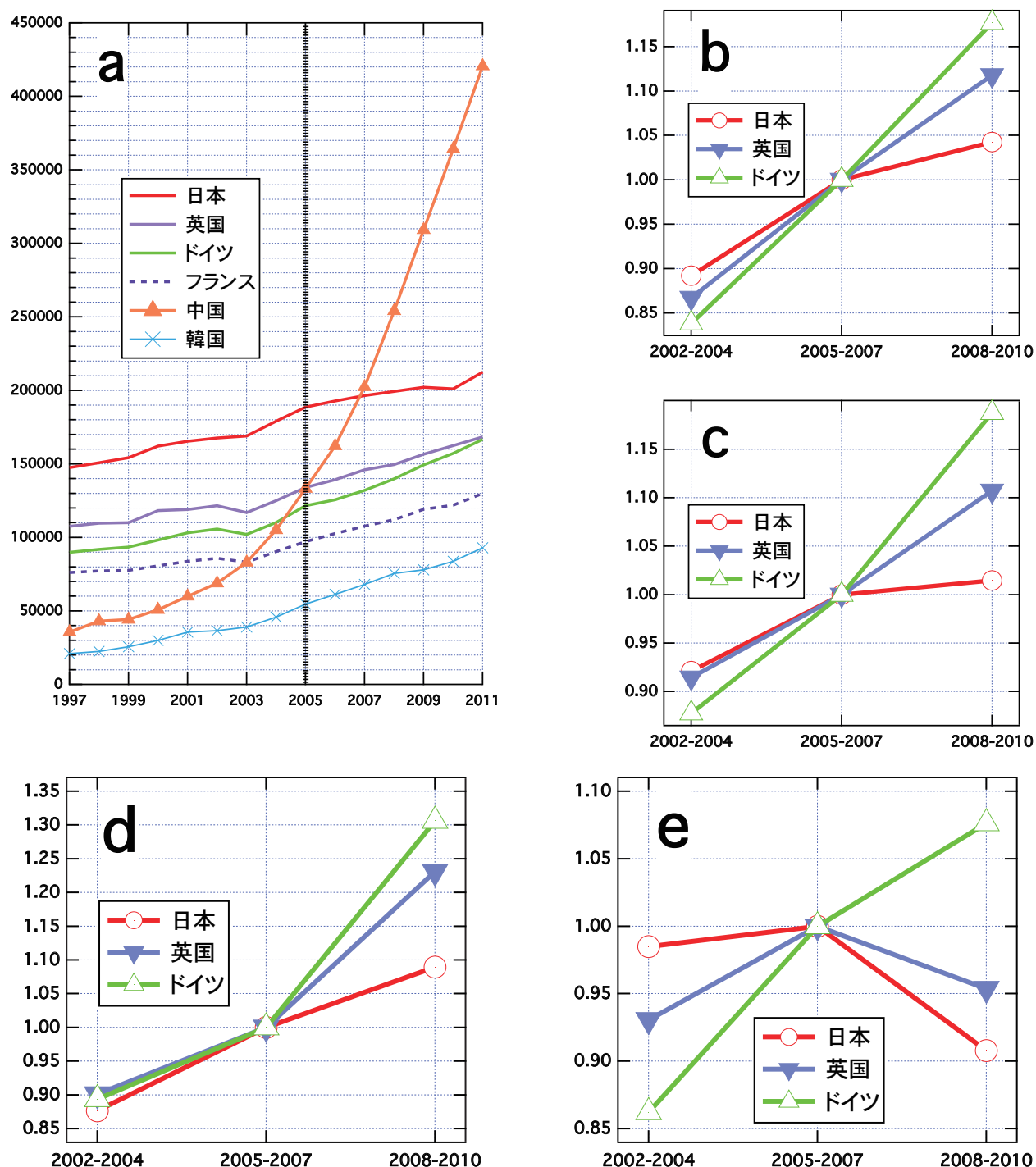


図 2 論文著者数の推移

(a:全ての著者数、以下 b 以降すべて 2005-2007 を 1 としたときの比
b:日英独の著者数の比、c:b のうち各年に初めて出現した著者数の比、
d:c のうち初出以降 2013 年までに一度も著者にならなかった人数の比、
e:c のうち初出以降 2013 年までにもう一度以上著者になった人数の比)

さらに、図 2c で示した「各年に初めて論文に名を連ねた著者」を「初出以降 2013 年までに一度も論文に名を連ねなかった著者」と「初出以降 2013 年までもう一度以上論文に名を連ねた著者」とに分解したものが図 2d と図 2e である。この図において、1 本目の論文から 2 本目の論文までの期間が長く二本目が未だデータベースに収録されていない著者の存在により、集計の構造上、どの国においても現在に近づくほど図 2d は増加し図 2e は減少する偏りを持つことに注意が必要である。図 2c、図 2d、図 2e から読み取れることとしては、A: 日本の論文生産が伸び悩んでいる要因のひとつとして、研究論文に名を連ねる著者の新規参入が伸び悩んでいることが関与している可能性がある、ということと、研究論文に新たに名を連ねた著者のうち、B1: その後も論文生産に関与する人数が減っている、もしくは B2: 1 本目の論文に著者として名を連ねて以後 2 本目の論文への関与までの期間が長期化していることが示唆されており、更なる分析が必要である。

精度評価

本研究では、ここまでに述べた、論文データベース内の論文著者 ID による研究者を単位とした計量書誌学的分析の信頼性を明らかにするために、今回利用した論文著者 ID である Scopus AuthorID の精度の評価も併せて行った。

評価方法は、科研費のデータベースである KAKEN の研究者番号を正解データとして扱い、KAKEN に成果報告として掲載されている論文書誌情報を媒介して研究者番号と AuthorID を接続し、以下の 2 つの精度を表す指標によって行った。

- ある 1 つの研究者番号に紐付いた AuthorID の数を論文数で重み付けして「再現率」を算出
 - ある 1 つの AuthorID に紐付いた研究者番号の数を論文数で重み付けして「適合率」を算出
- 評価結果を表に示す。論文書誌情報によって媒介された AuthorID は 75,405 ID、研究者番号は 69,017 番号であった。これを母集団としたとき、平均再現率は 98.44%、平均適合率は 99.36%であった。

表 1 Scopus 内著者 ID の精度評価結果

	全体	論文 2 報以上のみ
AuthorID の数	75,405	54,948
研究者番号の数	69,017	56,038
平均再現率	98.44 %	98.09 %
平均適合率	99.36 %	99.14 %

以上のように、日本の研究者に関しては実用的な精度であることを確認した。しかしながら、本検証の限界あるいは留保事項として以下の 3 点が存在する。

1. 日本の研究者のみを対象にした検証である
2. 正解データとして使用している KAKEN に非掲載の研究・研究者は対象外である
3. 2 と同じ理由により、KAKEN 自体の名寄せ精度に検証結果が依存する

特に 1 に関しては、各国の姓名システムの違いに大きく依存する可能性がある。主要国の研究者に関して、同姓同名別 AuthorID が存在する比率を調べると、日本:62.2%、米国:33.6%、英国:33.8%、ドイツ:26.8%、フランス:20.5%、中国:67.9%、韓国:60.0%と、国もしくは言語によって大きく異なる（日本人の同姓同名存在率に関しては、Yamauchi et al. (2014) が電話帳データを用いて、漢字表記では 67.7%と算出している）。論文データベース内の同姓同名別 AuthorID の数は ID の精度そのものに依存することから、直ちに日本と米欧の名寄せの困難さを比較する変数として用いることはできないが、いずれにしても各国間に精度の違いが存在する可能性は否定できない。

Isamu Yamauchi, Koichiro Onishi, and Takamasa Suzuki (2014) "How to trace mobile Japanese inventors? A rare name approach with patent data", *IIPR Discussion Paper* 2014-001. [forthcoming]