

Title	学際研究の実施は研究者の生産性にどのように影響するか
Author(s)	清川, 朝日; 鈴木, 崇史; 芳鐘, 冬樹
Citation	年次学術大会講演要旨集, 25: 493-498
Issue Date	2010-10-09
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/9345
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

学際研究の実施は 研究者の生産性にどのように影響するか

○清川朝日(筑波大学), 鈴木崇史(東洋大学), 芳鐘冬樹(筑波大学大学院)

1 はじめに

今日, 様々な領域で学際的な研究が盛んになされており, これらには複雑化した問題への解決策の提供, 複数分野の協働による研究の高度化, さらに新領域の萌芽といった役割が期待されている. このように学際研究に関心が高まる一方, 学際研究そのものの性質を明らかにしようとする研究も行われている. 学際研究を研究活動一般の中でいかに扱い, 他との差別化を図ることは重要なテーマである. これまで学際研究の性質について検討した研究が行われてきたが, 質的なアプローチが多く(e.g., Porter et al., 2006), 量的な分析においても「論文」を対象とした引用分析が多く(e.g., Rafols & Meyer, 2010), それらを行った「研究者」に関する分析は少ない.

本研究の関連研究としては, 学際研究自体のインパクトを測定するものがある. 例えば, 学際性の高低とインパクトの高低が, 必ずしも正の相関を見せないことが報告されている(Rinia et al., 2002). また, Lariviere & Gingras (2010)は, ①学際性の高低とインパクトの相関の高低は分野によって異なる, ②学際性の高低が極端な雑誌は, インパクトが低い, の二点を示した. これらは被引用回数やインパクトファクターといった指標を用いた研究であり, 学際研究の性質を計量的に明らかにする試みである. しかしながら, 本研究のように学際研究と生産性との関連を検討した研究は見られない.

本研究では学際研究の研究者への影響, 中でも生産性への影響に焦点を当て, 計量的手法を通して明らかにすることを試みている. 本研究では, **Environmental Sciences** 分野の論文の研究者を対象に, 研究者を学際研究実施群・非実施群に分け, それぞれについて計量された変数を用いた比較・分析を行った. さらに, それらを二期間(前者は学際研究実施の前後)に分けて分析を行うことで, 明確に実施後の影響について検討し, 学際研究実施の影響に関する知見を得ることを目指した.

2 分析手法

2.1 Random Forests

本研究では, 機械学習法の一つである Random Forests (Breiman, 2001)による判別・回帰分析を試みた. Random Forests により, 与えられたデータの分類・回帰分析とそれらに係る変数の重要度が測定される. 本研究ではこれを利用し, 学際研究実施の有無の判別を中心とした分析を行った. もし実施群・非実施群で分類が可能であるとすれば, それがどのような変数に基づいて行われうるかが検証できると推測される. Random Forests は, 判別・回帰と同時にそれらに係る変数の重要度を計算するため, 本研究で採用した. Random Forests を用いて次の観点を明らかにする. すなわち:

1. 学際研究の実施群・非実施群に見られるデータの違いは何か. また, 違いがあるとすれば, それが何によってもたらされるのか.
2. 学際研究の実施前後で研究者のデータに違いは現れるか. また, 非実施の研究者との間に差異は見られるか.
3. 学際研究の実施群・非実施群とで生産性に影響する要因は異なるかどうか.

2.2 変数

研究者の分析に用いる変数として, 対象年数(Y), 論文数(完全計数法: PC , 第一著者計数法: PF), 収載誌数(NJ), 共著者数(CA), 実施までの経過年(O)¹を基本とする. Random Forests を用いた分析では, ①(Y)と(O)以外の変数は自然対数を取り, ②後述の年数の差による影響を補正するため, 各変

¹非実施群においては, データに現れた著者の最初の論文の発行年と, 抽出された元論文の発行年との差を用いる.

数を年数で割った値を用いる(記号の末尾に $_{ley}$ を付与する). **Random Forests** で用いる変数の一覧を表 1 に示す. 以下の分析では, それぞれに用いられるデータ全件の半数を訓練用データとし, 全件を対象として分析した.

Table 1 Random Forests で用いる変数と略記号一覧

	略記号	前期 (A)	後期 (B)	ログ/年数	平均
論文数(完全計数法)	PC	APC	BPC	PC_{ley}	PC_{av}
論文数(第一著者計数法)	PF	APF	BPF	PF_{ley}	PF_{lav}
収載誌数	NJ	ANJ	BNJ	NJ_{ley}	NJ_{av}
共著者数	CA	ACA	BCA	CA_{ley}	CA_{av}
実施までの経過年	C	-	-	-	-

3 データ

情報源として, Elsevier 社の提供する引用索引データベース **Scopus**²を用いる. 対象は原著論文に限定した. 以下の手順に従い, 実施群・非実施群に分けてデータを収集した.

- 実施群
 1. キーワード検索により「学際研究」である論文を特定する
 2. 1.をもとに著者の一覧を取得し, 著者 ID 検索を行い, データをダウンロードする
- 非実施群
 1. 2003 年の論文から「学際研究」を除外し, 表示された順に論文 100 件を特定する
 2. 1.をもとに著者の一覧を取得し, 著者 ID 検索を行い, データをダウンロードする

データの抽出においてはその方法が問題となり, 分析方法によっては, 例えば, ある学際的な研究分野の全ての文献を対象とする研究もあるが, 本研究では, 学際研究をキーワード検索により特定する³. **Scopus** では, 以下の二種類のキーワードが与えられ⁴, これにより精度の高い検索が可能である.

- Author Keywords :
論文の著者が論文に対して付与するキーワード. 論文に記載されているものを検索可能としたものの.
- Index Keywords :
文献が **Scopus** に登録される際に付与されるキーワード. これらには **Scopus** 外のデータベースの索引語(MEDLINE の MeSH など)が用いられ, 抄録とインデクシングデータベースを元に人手で付与される場合がある.

各著者について, 学際研究を実施した直前の 5 年間の前期, 実施した年以降の 5 年間の後期として, それぞれの範囲内にある論文を特定し, 各変数を計量した. 表 2 に, 取得したデータの基本的数量として, 元となった論文数(OP), 分析対象の著者数(A), 分析対象となった著者の期間内の論文数(NP)⁵, 分析の対象となった年数の平均(Y_{av})を示す. 実施群・非実施群で(Y_{av})の差が大きいのは, 実施群として検出された論文の発行年に, 比較的最近のものが多かったためである⁶. **Random Forests** を用いた分析では, (Y)と(O)を除く全ての変数で自然対数を用い, 年数の差による影響を考慮しこれらを年数で割った値を用いる.

² Elsevier B.V. Scopus. <http://www.scopus.com/>, (accessed 2010-09-04).

³ ここで懸念されるのは, 「学際研究」というキーワードを付与しながらも, 必ずしも学際的でない研究(例えば, 本研究のような「学際研究」に関する研究)が抽出される場合であるが, これらは, 1. 社会科学系の雑誌に中心的に掲載され, 2. 様々な分野の雑誌にも掲載されるが, 分散しているため, 大きな影響は及ぼさないと考えられる.

⁴ Elsevier B.V. Coverage of metadata. Sciverse.

<http://www.info.sciverse.com/scopus/scopus-in-detail/content-coverage-guide/metadata/>, (accessed 2010-09-04).

⁵ 対象となった異なる著者が同一の文献の著者として現れる場合があるが, その重複は除外した.

⁶ 例えば, 2008 年に学際研究を実施した著者は, 2010 年現在で最大値が 3 年にしかなりえない. また, 「学際研究」をキーワードに持つ研究は近年増加傾向にあり, これも学際研究実施後の年数の少ない著者が相対的に多くなる原因と考えられる.

Table 2 データの基本的数量

	<i>OP</i>	<i>A</i>	<i>NP</i>	<i>Y_{av}</i>
実施群	74	221	3174	4.80
非実施群	100	361	6718	7.33
計	174	582	9892	6.37

4 分析結果

4.1 学際研究実施・非実施の判別

判別の結果を表 3 に、判別に係る変数の重要度を図 1 に示す。実施群の誤判別率はやや高く、2 割程度となっているが、全体としては 1 割程度の誤判別率であり、実施群・非実施群による違いが判別結果に現れていると考えられる。変数の重要度に目を移すと、前期よりも後期の変数が上位に位置している。中でも(*BCA_{lev}*)が大きく影響しており、これは後述の分析においても変数として重要度が高くなっている。*(BNJ_{lev})*、*(BPC_{lev})*の変数も上位に位置しており興味深い。以上より、実施群・非実施群の著者では前期・後期の変数により異なる傾向を見せると考えられる、続く分析ではこれらが具体的にどのような違いを生み出しているかを検討する。

4.2 前期・後期の判別

著者のデータを前期と後期に分け、それらの情報を元に前期・後期の判別を試みた。これにより、判別に係る重要度の高い変数は、学際研究の影響が大きいと推測できる。表 4 に前期・後期別に見た数量を、表 5 に判別結果と誤判別率を示す。非実施群の誤判別率の方が実施群よりも 5% 程度高いが、全体として 2 割程度となっている。この結果から、実施群・非実施群どちらにおいても前期・後期の傾向に違いがあると考えられる。図 2、3 にそれぞれの判別に係る変数の重要度を示した。*(PC_{lev})*は両群で 2 位以上にあり、また実施群においては 3 位であった(*CA_{lev}*)が、非実施群においては最も重要な変数として位置づけられている。さらに、非実施群では変数間で重要度の差がある程度開いているのに対して、実施群では(*PF_{lev}*)を除く 3 変数が拮抗している。学際研究実施群の研究者は、生産性だけでなくより多様な側面により影響を受けていることが示唆される。

4.3 生産性の回帰

最後に、学際研究の実施による生産性への影響について検討するため、実施群・非実施群それぞれの(*PC_{lev}*)、つまり全期間の生産性の変数を従属変数として回帰分析し、重要な変数を特定した。生産性に関わる 3 変数(*PC_{lev}*、*APC_{lev}*、*BPC_{lev}*)以外の変数を独立変数として投入した。変数の説明率は、実施群で 5 割程度、非実施群で 6 割程度であった。実施群・非実施群それぞれにおける生産性の回帰結果を図 4、5 に、各変数の重要度を図 6、7 に示す。実施群・非実施群において各変数の重要性の差異が観察された。実施群では(*BCA_{lev}*)および(*BNJ_{lev}*)の影響が特に大きく、非実施群では前期・後期の指標ともに上位にあり、変数間の重要度の差が実施群よりも小さくなっている。この結果から、学際研究の実施による研究者への傾向に影響があるとすれば、共著者数と収載誌数への影響が特に強いと推測される。

5 おわりに

本研究では、学際研究の実施が研究者の生産性にどのように影響するか検討するために分析を行った。実施群・非実施群の分析結果からは、実施前後の変数に関して傾向の差異が見られた。学際研究の実施は、一般に重要視される論文生産性だけでなく、共著者数・収載誌数の広がりといった側面への影響が強いと考えられる。また、実施・非実施の判別および生産性の回帰において、前期の指標は有効性が低い結果となったことも、学際研究の実施による影響を見出す手がかりとなろう。

今回は試行的な研究であったため、Environmental Sciences 分野の研究者のみを対象とし、分析の範囲を狭めた。今後は分野を増やすとともに、個々の変数に対する解釈を深めていきたい。また、重要度の高いことが示唆された共著者に関係する変数である、所属機関数や著者の役割に関連した指標にも着目したい。

参考文献

- 1) Breiman, L. (2001) "Random forests," Machine Learning, Vol. 45, pp. 5-32.
- 2) Lariviere, V., Gingras, Y. (2010) "On the relationship between interdisciplinarity and scientific impact," Journal of the American Society for Information Science and Technology, Vol. 61, No. 1, pp. 126-131.
- 3) Porter, A.L., Roessner, J.D., Cohen, A.S., & Perreault, M. (2006) "Interdisciplinary research: meaning, metrics and nurture," Research Evaluation, Vol. 15, No. 3, pp. 187-195.
- 4) Rafols, I., Meyer, M. (2010) "Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience," Scientometrics, Vol. 82, No. 2, pp. 263-287.
- 5) Rinia, E.J., van Leeuwen, T.N., & van Raan, A.F.J. (2002) "Impact measures of interdisciplinary research in physics," Scientometrics, Vol. 53, No. 2, pp. 241-248.

Table 3 Confusion Matrix : 実施群・非実施群の判別結果

	実施群	非実施群	計	誤判別率
実施群	179	42	221	19.00%
非実施群	35	326	361	9.70%
計			582	13.23%

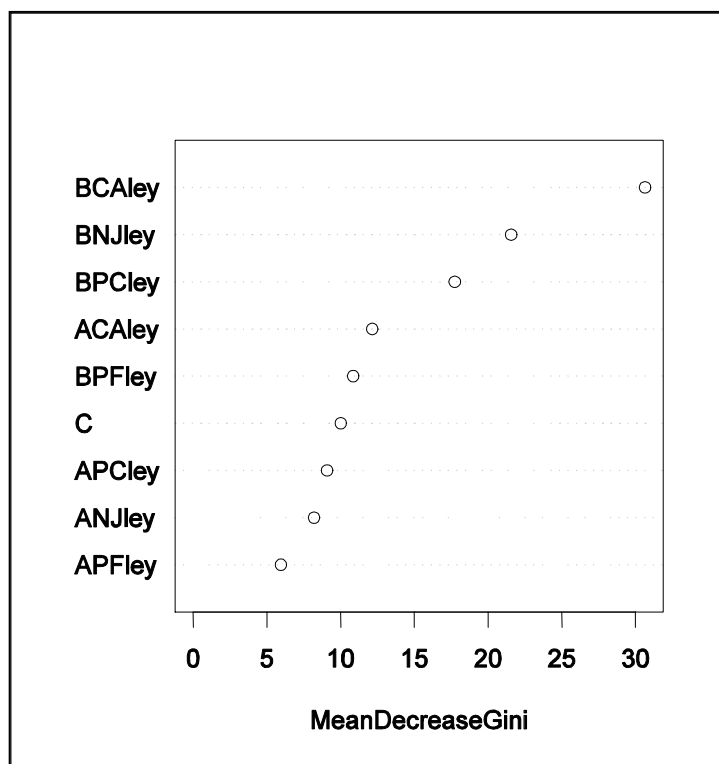


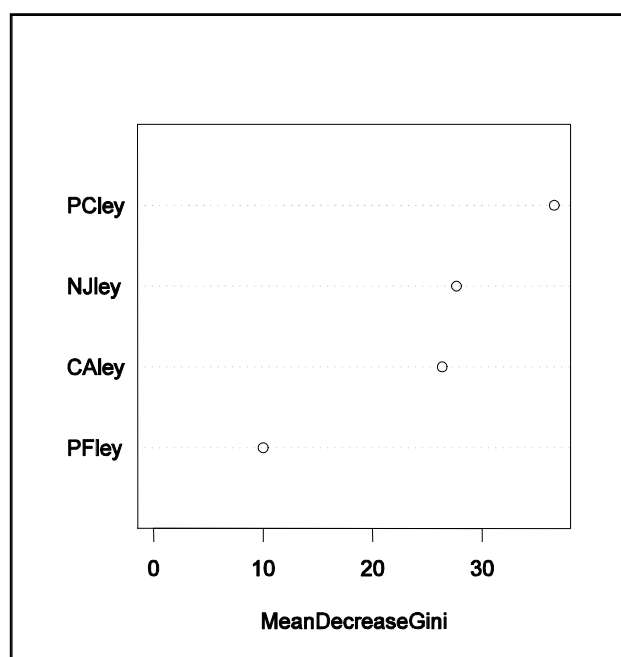
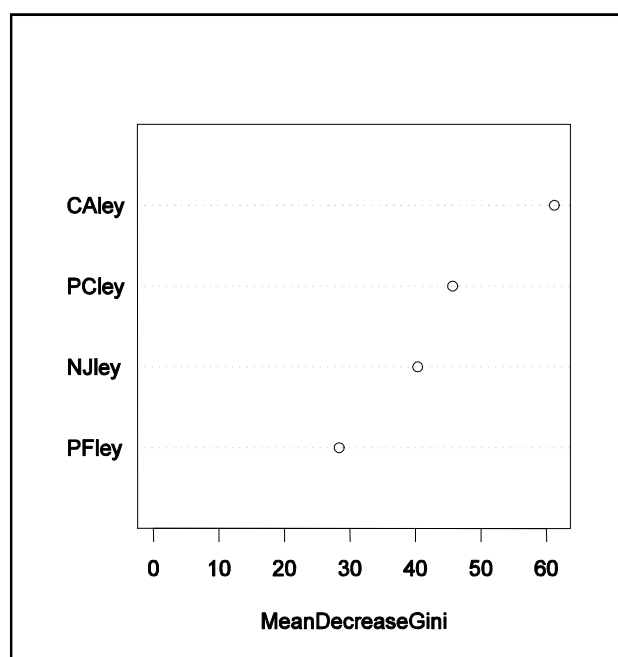
Figure 1 実施群・非実施群の判別に係る変数の重要度

Table 4 前期・後期別に見たデータの数量

	期間	A	Y_{av}	PC_{av}	PF_{av}	CA_{av}	NJ_{av}
実施群	前期	221	2.54	6.59	1.86	18.68	4.43
	後期		2.25	7.77	1.95	23.42	4.86
非実施群	前期	361	3.35	7.52	2.08	18.87	4.84
	後期		3.98	11.09	2.24	32.03	7.20

Table 5 Confusion Matrix : 前期・後期の判別結果

	期間	前期	後期	計	誤判別率
実施群	前期	186	35	221	15.84%
	後期	31	190	221	14.03%
				442	14.93%
非実施群	前期	297	64	361	17.73%
	後期	83	278	361	22.99%
				722	20.36%

Figure 2 実施群における前期・後期の判別に
係る変数の重要度Figure 3 非実施群における前期・後期の判別に
係る変数の重要度

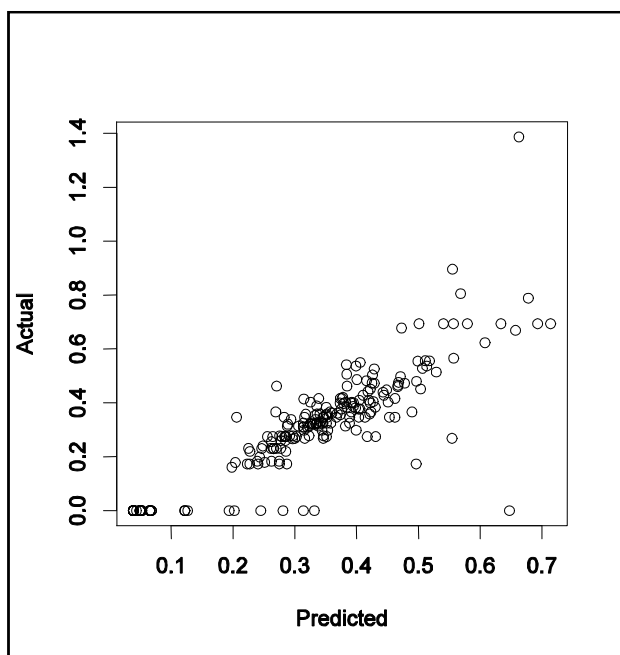


Figure 4 実施群における生産性の回帰結果

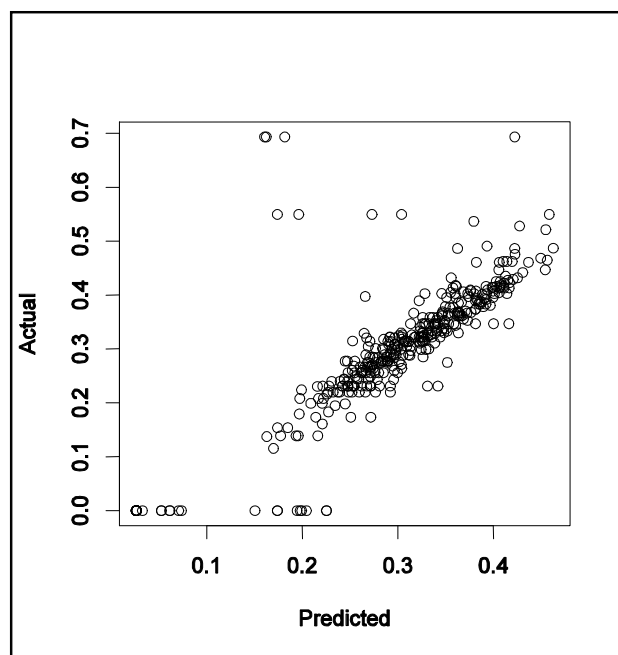


Figure 6 非実施群における生産性の回帰結果

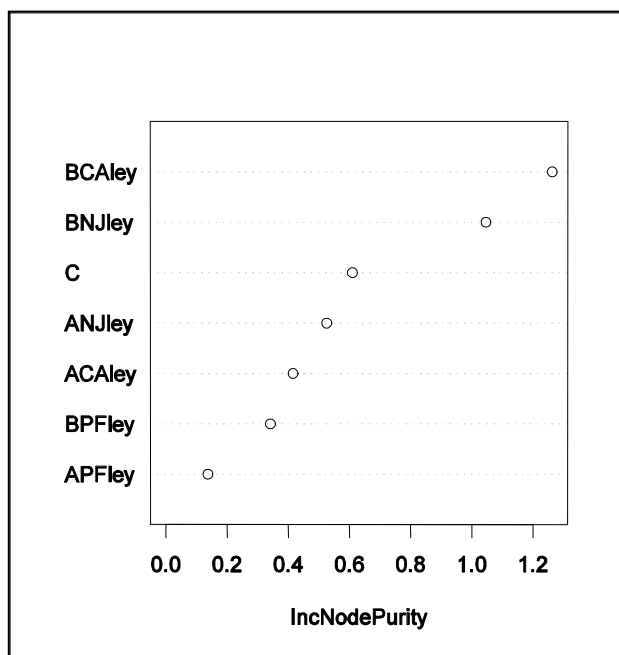


Figure 5 実施群における生産性の回帰に係る変数の重要度

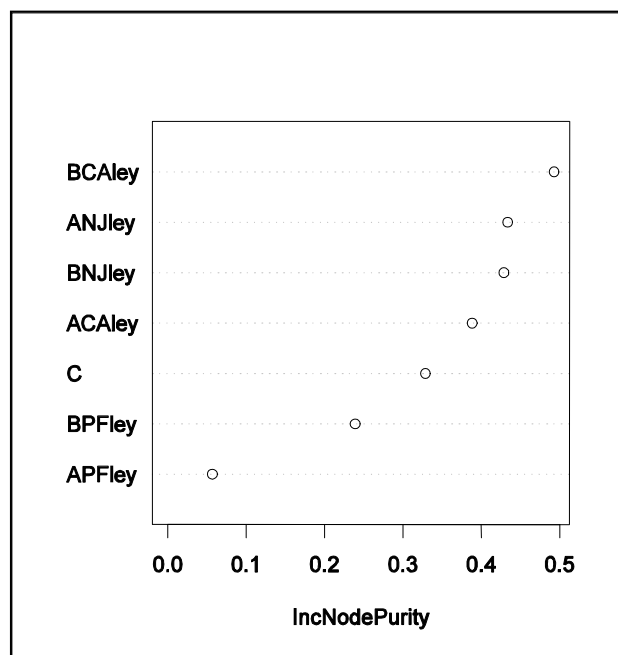


Figure 7 非実施群における生産性の回帰に係る変数の重要度