

論文データを用いた著者の貢献度推定手法の評価

渡辺 曜大^{†1} 市瀬 龍太郎^{†2}

本論文では、論文データを用いた著者の貢献度推定手法の評価結果について報告する。

Evaluation of a Method for Contribution Estimation of Authors

YODAI WATANABE^{†1} and RYUTARO ICHISE^{†2}

This paper is a report on an evaluation of a method for the contribution estimation of authors.

1. はじめに

近年になり、学術研究の細分化が進んでいる。それに伴い、学術研究においては、多くの知見を統合しながら研究を進めなければならなくなっている。Jones らの研究¹⁾によると、科学と工学の分野において、1975年には、単著の論文が3割程度であったのに対して、2005年には、1割程度にまで低下している。その一方で、複数の大学にまたがる多様な研究者で、研究チームを構成して書かれた論文が3割以上に増えてきている。このような現象は、科学と工学の分野のみならず、社会科学の分野でも見られ、多様な知の統合が、新たな知を生み出すのに欠かせなくなっていると言えるであろう。

たくさんの研究者がお互いに知を出し合い、研究をすれば、常に新しく画期的な研究ができるわけではない。どのような研究者がチームを組めば、新たな知が生み出されるのかを予

測するために、リンク予測を用いた研究などが行われてきている²⁾。そのような研究では、研究者の論文に対する貢献度は、等価なものとして取り扱われている。しかし、実際には、論文に対して、研究者が貢献する部分は異なっているであろう。例えば、病気の治療に対して、新たなデバイスを用いる論文であったならば、医学の専門家と工学の専門家の共同研究になるであろう。しかし、もし、その論文がデバイスの作成方法に重点をおいた研究であるならば、論文そのものに対する貢献度としては、医学の専門家は相対的に貢献度が低く、工学の専門家が高いであろう。一方、もし、その論文が、治療に重点が置かれた論文であれば、逆に、医学の専門家の方が貢献度が高くなるであろう。したがって、各研究者が、共同研究をする時に、どのようなことに貢献できるのかを考える際には、それぞれの論文に対して、どのくらいの貢献をしたのかの情報も同定する必要が出てくる。そこで、本研究では、論文に対する著者の貢献度を推定する手法を提案する。

提案手法は、著者貢献モデルと呼ばれる確率モデルに基づいており、変分ベイズ法を用いることで、論文著者の貢献度を推定することができる。本論文では、さらに、人工データと実際の論文執筆データを用いて、このモデルの検証を行った結果についても報告する。まず、第2章で関連研究について述べ、本研究の特徴を明かにする。次の第3章で、本論文で提案する著者貢献モデルについて述べる。第4章では、著者貢献モデルに対して行った実験について述べ、提案手法の特徴について、議論を行う。最後に、第5章で本研究をまとめる。

2. 関連研究

論文から、確率モデルを用いて著者と論文の関係を推定する研究は、これまでもいくつかの研究が行われている。著者-話題モデル³⁾では、論文の情報から、著者がどの話題に関連しているかを示す分布を計算するための確率モデルを提案している。このモデルでは、著者、話題、語の間の関係をモデル化し、論文の執筆者の情報と語の情報から、著者と話題の関係を推定する。著者-ペルソナ-話題モデル⁴⁾では、著者が複数の話題分布を持つことができるようにすることで、著者-話題モデルを拡張している。また、潜在興味話題モデル⁵⁾では、著者の話題分布間に類似性を導入した確率モデルを提案している。このように、著者と話題の間の関係を確率モデルを用いて推定する方法はいくつか研究されてきているが、これらの研究は、本研究で提案するような、著者の貢献度を直接推定するような確率モデルでは無い。

一方、ある論文の他の論文に対する貢献度を確率モデルによって推定する研究が Dietz らによってなされている⁶⁾。この研究では、論文が引用論文からの影響によって構成されると

^{†1} 会津大学

University of Aizu

^{†2} 国立情報学研究所

National Institute of Informatics

いう確率モデルや、それに加えて、論文は引用文献にない革新的な視点を付加しているという確率モデルを提案し、論文の貢献度を推定する。また、Mann らの研究⁷⁾では、話題モデルを用いることで、話題多様性などの論文同士への影響度を測る手法を提案している。これらの研究は、確率モデルを用いて、論文間の貢献度を推定する手法とみなすことができるが、本研究で取り組む、著者の論文への直接的な貢献度を測るような研究とは異なると言える。

3. 著者貢献モデル

3.1 論文の生成モデル

本論文で提案する著者貢献モデルにおいて、論文の生成モデルは次のようになっている。

- 文書 d の著者集合 A_d に対して、著者選択（貢献度）の多項分布 ψ_d から、著者 a を選択。
- 選択した著者 a に応じて、トピック選択の多項分布 θ_a から、トピック z を選択
- 選択したトピック z に応じて、単語選択の多項分布 ϕ_z から、単語 w を選択

この時、観測可能なのは、文書 d の著者集合 A_d と、文書中の単語 w である。

3.2 変分ベイズ法による推定

本研究では、パラメータ推定に変分ベイズ法⁸⁾を用いる。変分ベイズ法では、潜在変数 Z およびパラメータ θ のテスト分布 $q(Z)$, $q(\theta)$ を導入し、適当な初期分布を仮定して以下の2つのステップを収束するまで繰り返すことによって真の事後分布 $p(Z, \theta|D)$ を近似する。

- VB-E ステップ

$$q(Z) \leftarrow C \exp\langle \log p(D, Z|\theta) \rangle_{q(\theta)} \quad (1)$$

- VB-M ステップ

$$q(\theta_i) \leftarrow C p(\theta_i) \exp\langle \log p(D, Z|\theta) \rangle_{q(Z)q(\theta_{-i})} \quad (2)$$

ただし、 C は規格化定数であり、 θ_{-i} は θ_i 以外の θ の成分をあらわしている。

前節の確率モデルでは、 $D = (\mathbf{d}, \mathbf{w})$, $Z = (\mathbf{a}, \mathbf{z})$ であり、

$$p(D, Z|\theta) = \prod_i p(d_i)p(a_i|d_i)p(z_i|a_i)p(w_i|z_i)$$

とあらわされる。多項分布の自然共役事前分布は Dirichlet 分布だから、事前分布を

$$p(a|d) \sim \text{Dir}(\theta_{da}^0), \quad p(z|a) \sim \text{Dir}(\theta_{az}^0), \quad p(w|z) \sim \text{Dir}(\theta_{zw}^0)$$

とおけば、事後分布もパラメータ $\theta_{da}, \theta_{az}, \theta_{zw}$ をもちいて

$$q(a|d) \sim \text{Dir}(\theta_{da}), \quad q(z|a) \sim \text{Dir}(\theta_{az}), \quad q(w|z) \sim \text{Dir}(\theta_{zw})$$

とあらわすことができる。これらを (1) 式に代入して計算すると、潜在変数 \mathbf{a}, \mathbf{z} のテスト分布 $q(\mathbf{a}, \mathbf{z})$ は

$$q(\mathbf{a}, \mathbf{z}) = \prod_i q_i(a_i, z_i)$$

と因数分解され、分布 q_i は d_i, w_i のみに依存することがわかる。これを q_{dw} とあらわせば、 q_{dw} の更新則は

$$q_{dw}(a, z) \leftarrow C \exp(\psi(\theta_{da}) - \psi(\theta_d) + \psi(\theta_{az}) - \psi(\theta_a) + \psi(\theta_{zw}) - \psi(\theta_z))$$

であたえられる。ただし、 C は正規化定数、 ψ は Digamma 関数をあらわし、

$$\theta_d = \sum_a \theta_{da}, \quad \theta_a = \sum_z \theta_{az}, \quad \theta_z = \sum_w \theta_{zw}$$

とおいた。一方、(2) 式より、パラメータ $\theta_{da}, \theta_{az}, \theta_{zw}$ の更新則は、論文 d における単語 w の頻度を F_{dw} とするとき、

$$\theta_{da} \leftarrow \theta_{da}^0 + \sum_{z,w} F_{dw} q_{dw}(a, z) \quad (3)$$

$$\theta_{az} \leftarrow \theta_{az}^0 + \sum_{d,w} F_{dw} q_{dw}(a, z) \quad (4)$$

$$\theta_{zw} \leftarrow \theta_{zw}^0 + \sum_{d,a} F_{dw} q_{dw}(a, z) \quad (5)$$

と求まる。アルゴリズムが収束してパラメータ θ_{da} の値が求まったら、貢献度 $\text{con}_d(a)$ は

$$\text{con}_d(a) = E[q(a|d)] = \frac{\theta_{da}}{\theta_d}$$

と計算できる。

なお、ここで得られた更新則はトピック z の置換に関して不変であることに注意しておく。したがって、初期値 θ_{az}^0 として一様分布を選ぶとすべてのステップにおいて θ_{az} も一様分布となり、正しい推定を行うことができない。そこで本研究では、初期分布の各パラメータに対し 0 に収束する乱数を付加してアルゴリズムを実行した。この修正により、適当な条件の下で十分な精度をもってパラメータを推定できることが、以下に記す人工データに対する実験によって確認することができる。

4. 実験

4.1 人工データに対する実験

上記の変分ベイズアルゴリズムを用いて実際の論文執筆データを扱う前に、まずこれを人工的に生成されたデータに対して適用し、その推定誤差を調べた。ここで、人工データは以下の要領で生成した。

表 1 人工データに対する実験結果

n_d	n_a	n_z	n_w	t_w	誤差
1000	100	10	100	1000	0.035
1000	100	10	100	100	0.164
100	100	10	100	1000	0.253
1000	100	10	1000	1000	0.033
1000	100	50	100	1000	0.040

- 論文数 n_d , 著者数 n_a , トピック数 n_z , 単語数 n_w を定める.
 - 各論文の著者数, 各著者のトピック数, 各トピックの単語数を適当な分布を仮定してランダムに定める.
 - 各論文の著者分布, 各著者のトピック分布, 各トピックの単語分布を適当な分布を仮定してランダムに定める.
 - 一論文の延べ単語数 t_w を定め, 上記確率モデルに従って各論文の単語を生成する.
- 実験の設定値とその結果を表 1 に示す. ただし, 真の貢献度分布と推定貢献度分布の間の変動距離をすべての論文について平均したものを誤差としている.

この実験結果から, 以下のことがわかる.

- n_a を固定した際に, n_d, t_w が大きい方が誤差が小さくなる. 特に, n_d, t_w が十分に大きければ, 誤差は十分小さくなる
- n_d と t_w では, n_d を大きくした方が誤差を小さくするのに有効である

4.2 論文執筆データ

本研究では, 論文データベース Scopus^{*1}から, 実際のデータを取得して, 提案手法の検証を行った. データの作成に当たり, 市瀬らガリサーチマイニングで用いた手法⁹⁾と同様な手順を踏んだ. まず, データの作成のために, さまざまな Web サイトなどを参考に, 計算機科学の分計算機科学の分野の 14 の研究分野 (計算機ネットワーク, ヒューマンコンピュータインタラクションなど) を決めた. 次に, それぞれの研究分野に対して, Scopus から近年多数の論文を執筆しているアクティブな著者を約 100 人ずつ抽出した. 次に, その著者が 2000 年から 2007 年までの間に執筆した論文 14,238 本を抽出した. この論文に出現する共著者を含めた著者の数は 17,515 名となった. また, 延べ著者数は, 45,667 名である. 次に, これらの論文に対して, Scopus のデータベースから, それぞれの論文の, 概要, 著者

*1 <http://www.scopus.com/>

表 2 実験に使ったデータの統計値

論文数	14,238
著者数	17,515
延べ著者数	45,667
単語種類数	19,306
延べ単語数	1,164,018

キーワード, 索引キーワードを取り出し, その中に含まれる名詞を抽出した. そして, ステミングを実施した結果を, その論文のキーワードとした. その結果, 単語の種類が 19,306 個, 延べ単語数が 1,164,018 個となった. これらの実験に使ったデータの統計値を表 2 にまとめた.

以下, 論文執筆データに関する記号をいくつか導入し, 代表的な著者集合を定義する. 論文 d の著者 $a \in A_d$ の順番を $\text{ord}_d(a)$, 貢献度を $\text{con}_d(a)$ とあらわせば, d の第一著者, 最終著者の集合 A_d^F, A_d^L はそれぞれ

$$A_d^F = \{a \in A_d | 1 = \text{ord}_d(a) < |A_d|\}, \quad A_d^L = \{a \in A_d | 1 < \text{ord}_d(a) = |A_d|\}$$

とあらわされ, 貢献度が最も高い著者の集合 A_d^+ は

$$A_d^+ = \{a \in A_d | \text{con}_d(a) = \max_{a \in A_d} \text{con}_d(a)\}$$

とあらわされる. ここで, 第一著者, 最終著者は著者数 $|A_d|$ が 2 以上の論文に対して定義されていることに注意しておく. また, 論文数 i の著者の集合を $A(i)$ とあらわす. すなわち,

$$A(i) = \{a | \sum_{d: a \in A_d} 1 = i\}$$

と置く.

4.3 スコア

すでに述べたとおり, 本研究で用いた変分ベイズアルゴリズムでは初期分布のパラメータに乱数を用いている. 論文執筆データに対して同一条件で実験を複数回実行してみたところ, 得られた貢献度分布は実験ごとのばらつきが大きいことがわかった. そこで本研究では, 個々の論文の貢献度分布には注目せず, 以下で与えるスコアの統計を調べることにする.

いま仮に「第一著者をもっとも貢献度が高い」と仮定して, 推定した貢献度分布からこの仮定が正しいか否かを検証する状況を考えることにする. そのために例えば, 第一著者の貢献度の推定値が最大となっている論文の割合を計算し, これが 0.3 と求めたとしよう. 問題は, この 0.3 が大きい小さいかをいかに判定するかである. ここで, もし貢献度最大の著者が論文中の順番とは独立にランダムに決まるとすると, 第一著者が最大となる論文の割

合の平均値は

$$\sum_d \frac{1}{|A_d|}$$

とあらわされるから、この平均値を基準として 0.3 がこれよりも有意に大きければ「第一著者はもっとも貢献度が高い」と言えることになる。

本研究では、第一著者に限らずさまざまな著者集合の貢献度について調べるため、基準となる平均値が何か普遍的な値をとると都合がよい。そこで、この基準（平均値）が 1 となるようにスコアを定める。具体的には、論文 d の推定貢献度最大の著者に d の著者数 $|A_d|$ を得点として付与することにする。ただし、推定貢献度最大の著者が複数存在する場合は、平均値が 1 となるように著者数 $|A_d|$ をその人数で割った値を貢献度最大の各著者に得点として付与するものとする。すなわち、論文 d の著者 $a \in A_d$ のスコア $\text{score}(d, a)$ を

$$\text{score}(d, a) = \begin{cases} \frac{|A_d|}{|A_d^+|} & \text{if } a \in A_d^+ \\ 0 & \text{otherwise} \end{cases}$$

と定めることにする。

以下、いくつかの著者集合についてこのスコアの平均を考える。そこで、関数の平均を以下のように定義しておく：集合 \mathcal{D} 上の実数値関数 f の集合 $\mathcal{X} \subset \mathcal{D}$ における平均 $E_{x \in \mathcal{X}}[f(x)]$ を、 \mathcal{X} 上の一様分布に従う確率変数 X を用いて

$$E_{x \in \mathcal{X}}[f(x)] = E_X[f(X)]$$

と定める。

4.4 トピック数とスコア

本研究の変分ベイズアルゴリズムを実行する上であらかじめトピック数 n_z を定めておかなければならない。第一著者、最終著者について、このトピック数 n_z を変えてスコアの平均

$$E_{d, a \in A_d^*}[\text{score}(d, a)]$$

($\omega \in \{F, L\}$) を計算した結果を図 1 に示す。なお、同一の条件で実験を 15 回行い、その分散をエラーバーとしてあらわしている。この図から、以下のことがわかる。まず、トピック数によらず第一著者のスコアは平均よりも低く、最終著者のスコアは平均よりも高い。さらに、(トピック数 2 を除いて) トピック数が大きくなるほどこの両者の平均からのずれも大きくなる。これは、トピック数を大きく選ぶと、単語分布の近似精度が上がるためと考えられる。しかし、同時に計算量も増えるため、大きなトピック数を用いて実験を行うことは計算時間の観点から望ましくない。そこで本研究では、論文執筆データの研究分野数 14 に合わせて、トピック数 $n_z = 14$ として以下の実験・解析を行う。

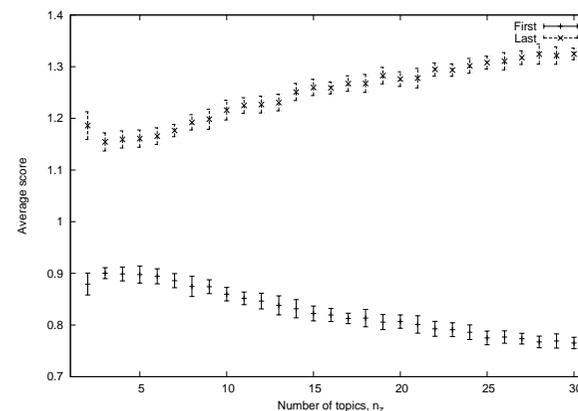


図 1 トピック数とスコア

4.5 著者の論文数と貢献度

いま論文 d の著者からなる集合 $A_d^* \subset A_d$ を一つ固定する。もし貢献度最大の著者集合 $A_d^+ (\neq \emptyset)$ が集合 A_d の置換に関して不変な確率分布に従って選ばれるならば、すなわち、任意の $a_d, a'_d \subset A_d$ に対して

$$|a_d| = |a'_d| \rightarrow \Pr[A_d^+ = a_d] = \Pr[A_d^+ = a'_d]$$

をみたすならば、

$$E_{A_d^+} \left[E_{a \in A_d^*} [\text{score}(d, a)] \right] = 1$$

が成り立つ。ここで $\text{score}(d, a)$ の値は有限だから、 A_d^+ が論文ごとに独立に選ばれれば、著者集合 A_d^* のスコアの平均について大数の法則が成り立つ。すなわち、論文数 n_d が十分に大きいとき高い確率で

$$E_{d, a \in A_d^*} [\text{score}(d, a)]$$

は 1 に近い値をとる。したがって、スコアの平均の 1 からのずれを、その著者集合が「平均的（標準的）な著者」と比較してどの程度論文に貢献しているかを調べるための指標とすることができる。

例えば、 $A_d^* = A_d \cap A(i)$ として論文数 i の著者のスコアの平均

$$\text{score}(i) = E_{d, a \in A_d \cap A(i)} [\text{score}(d, a)]$$

を計算すると図 2 を得る。著者の論文数が増えるにしたがって平均スコアも増え、論文数

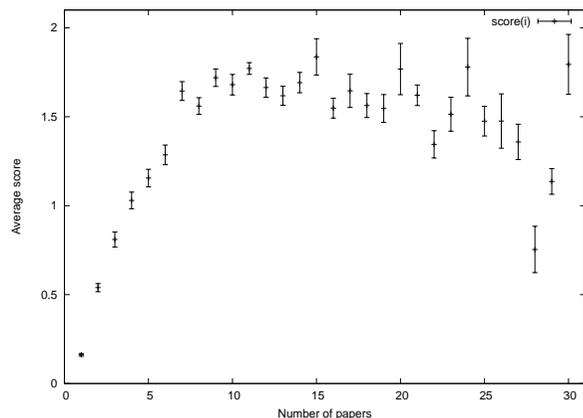


図 2 著者の論文数とスコア

10 程度で一定値に漸近している（あるいはその後減少に転じているかもしれない）。平均スコアが 1 となるのは論文数 4 本のあたりである。これはちょうど、「論文を 5 本程度書いて一人前」というわれわれの認識と一致している。

4.6 論文における著者の順番と貢献度

第一著者、最終著者の平均論文数はそれぞれ 7.4, 11.9 であり、論文数とスコアの関係（図 2）より第一著者のスコアが低く、最終著者のスコアが高くなることが予想される。したがって、図 1 の結果（第一著者のスコアが低く、最終著者のスコアが高い）が単に論文数の大小が反映されたものであるか否かを確認する必要がある。そこで、論文数 i の第一著者、最終著者のそれぞれについてスコアの平均

$$\text{score}^\omega(i) = \frac{E}{d, a \in A_d^\omega \cap A(i)} [\text{score}(d, a)]$$

($\omega \in \{F, L\}$) を計算して図 3 を得た。ここで、論文数 i の第一著者、最終著者の延べ数

$$n^F(i) = \sum_d |A_d^F \cap A(i)|, \quad n^L(i) = \sum_d |A_d^L \cap A(i)|$$

は表 3 で与えられる。

図 3 より、論文数同一の条件のもとでも、第一著者のスコアは最終著者のスコアに比べて低いことが分かる。多くの研究分野において成立していると考えられる「もっとも貢献度が高い著者を第一著者とする」という慣習に反する結果が得られたことになり、一見奇異に感じられるかもしれない。しかし、本研究で考える「貢献度」とは各著者のトピック分布を混

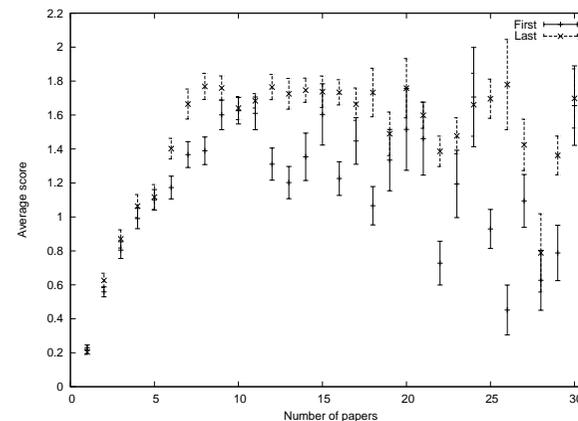


図 3 論文における著者の順番とスコア

表 3 論文数と著者数

i	$n^F(i)$	$n^L(i)$	i	$n^F(i)$	$n^L(i)$
1	3406	2125	16	234	390
2	2114	1109	17	148	330
3	1312	732	18	141	199
4	848	533	19	114	157
5	663	402	20	73	286
6	484	330	21	58	172
7	442	474	22	77	227
8	382	452	23	86	175
9	403	633	24	45	104
10	440	642	25	104	183
11	304	688	26	45	139
12	331	527	27	102	122
13	183	365	28	35	17
14	232	580	29	72	118
15	117	297	30	24	52

合する際の重みのことだから、これを「著者の持っているバックグラウンドが論文の内容にどの程度一致しているかの尺度」ととらえることにすれば、以下のように説明することができるだろう。

- 第一著者のスコアが低くなる理由

第一著者として積極的に論文を書く研究者は標準的な著者とは異なる固有の特徴（オリジナリティ）を持っていることが多く、共著論文のある種平均化された分布からみると標準的な著者に比べてずれが大きくなるため「貢献度」が低くなる。あるいは、第一著者は他の著者に比べて自分のバックグラウンドから外れた新しい研究分野に挑戦することが多い、と考えることもできる。

- 最終著者のスコアが高くなる理由

最終著者は、共同研究において指導的立場にある者になる場合が多い。したがって最終著者は、指導するグループが生産する論文の多くにおいて共著者となり、それ故そのグループにおいて標準的な分布をもつことになるが、標準的であるということは多くの共同研究の分布に近いということであり、これらの論文における「貢献度」も高くなる。

5. おわりに

本研究では、学術論文に対する著者の貢献度を推定する方法を提案した。著者の貢献度を取り入れた確率モデルを導入し、その上のパラメータ推定を行う変分ベイズアルゴリズムを導出した。このアルゴリズムをまず人工的に生成したデータに対して適用し、適当な条件のもとで推定誤差が十分に小さくなることを確認した。さらに、このアルゴリズムを用いて実データを解析するために著者の貢献度を反映したスコアを導入し、著者の論文における順番や論文数とこのスコアとの関係を調べた。その結果、

- 著者の論文数が4本のあたりで平均スコアが1となる

- 論文数同一の条件のもとで、第一著者のスコアは最終著者のスコアに比べて低い

ことがわかった。すでに述べたとおりこれらはもっともらしい結果であり、本研究の貢献度推定がある程度は成功していることを示唆していると考えられる。

最後に、今後の課題について述べる。まず、もっとも重要な課題の一つが、貢献度分布の推定誤差をさらに小さくすることである。今回は研究の第一歩ということで非常に単純な確率モデルを採用したが、このモデルを精緻化する必要があるだろう。また、さらに論文数・単語数の多い論文執筆データを用意することも有効であると考えられる。一方、本研究で得

られた実験結果に関して、著者の論文数を大きくしたときにスコアがある値に漸近するのがあるいは減少に転じるのかについて統計的に検証することも今後の課題である。もし減少に転じている場合は「責任をもって書ける論文の数には限りがある」ということを示唆しているようで興味深いし、あるいは論文執筆データによらず普遍的にある値に漸近するのであればその理由を考察することはさらに興味深い問題であると思われる。

参考文献

- 1) Jones, B.F., Wuchty, S. and Uzzi, B.: Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science, *Science*, Vol. 322, pp. 1259–1262 (2008).
- 2) Wohlfarth, T. and Ichise, R.: Semantic and Event-Based Approach for Link Prediction, in Yamaguchi, T. ed., *Proceedings of the 7th International Conference on Practical Aspects of Knowledge Management*, Vol. 5345 of *Lecture Notes in Computer Science*, pp. 50–61, Springer (2008).
- 3) Steyvers, M., Smyth, P., Rosen-Zvi, M. and Griffiths, T.: Probabilistic author-topic models for information discovery, in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 306–315 (2004).
- 4) Mimno, D. and McCallum, A.: Expertise Modeling for Matching Papers with Reviewers, in Berkhin, P., Caruana, R. and Wu, X. eds., *Proceedings of the 13th International Conference on Knowledge Discovery and Data Mining*, pp. 500–509, ACM (2007).
- 5) 川前徳章, 山田武士: 著者の興味と文書の内容の依存関係に着目した潜在変数モデル, *信学技報*, Vol. 109, No.51, pp. 19–24 (2009).
- 6) Dietz, L., Bickel, S. and Scheffer, T.: Unsupervised prediction of citation influences, in *Proceedings of the 24th international conference on Machine learning*, pp. 233–240, New York, NY, USA (2007), ACM.
- 7) Mann, G.S., Mimno, D. and McCallum, A.: Bibliometric impact measures leveraging topic analysis, in *JCDL '06: Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries*, pp. 65–74, New York, NY, USA (2006), ACM.
- 8) Attias, H.: Inferring Parameters and Structure of Latent Variable Models by Variational Bayes, in Laskey, K.B. and Prade, H. eds., *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence (UAI-99)*, pp. 21–30, S.F., Cal. (1999), Morgan Kaufmann Publishers.
- 9) Ichise, R., Fujita, S., Muraki, T. and Takeda, H.: Research Mining using the Relationships among Authors, Topics and Papers, in *Proceedings of the 11th International Conference on Information Visualization*, pp. 425–430 (2007).