様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　元 年　 ６ 月 １９ 日現在

機関番号：１３９０４
研究種目：若手研究(B)
研究期間：2016～2018
課題番号：１６Ｋ１６１５５

研究課題名（和文）ソーシャルメディアにおける学術文献言及量予測モデルの構築

研究課題名（英文）Analysis of research papers appearing in social media

研究代表者
吉田　光男（Yoshida, Mitsuo）

豊橋技術科学大学・工学（系）研究科（研究院）・助教

研究者番号：６０７３４９７８

交付決定額（研究期間全体）：（直接経費）　　2,400,000円

研究成果の概要（和文）：本研究では，新たな学術文献評価指標を開発することを目指し，ソーシャルメディアにおける論文の言及データを大規模に収集し，分野を横断した言及要因の分析を行った。その結果，理工学系の論文よりも人文社会学系の論文の方がソーシャルメディアでよく言及されることなどが明らかになった。また，ニュースの典拠論文を調査するために，ニュース記事に現れる学術雑誌名を自動抽出するアルゴリズムを開発し，高精度に抽出できることを確認した。

研究成果の学術的意義や社会的意義
従来の研究評価指標は理工系に有利であるとの指摘があり，理工学系の論文よりも人文社会学系の論文の方がソーシャルメディアで言及されるという結果は，従来の研究評価に加えてソーシャルメディアのデータを利用することで，研究評価における分野の有利不利を補正できる可能性を示している。また，ニュース記事から自動的に学術雑誌名を抽出できることは，そのニュースの典拠となる学術論文へのアクセスを提供し，多様な視点での理解を助けることに繋がる。

研究成果の概要（英文）：We aimed to develop a new academic research evaluation index. We therefore collected data on the mention of research papers in social media on a large scale, and analyzed factors of the mention. As a result, we found that the papers of the humanity and social science is more often mentioned to by social media than the papers of the science and engineering. In addition, in order to analyze the source papers of science news articles, we developed an algorithm to automatically extract the journal names appearing in news articles.

研究分野：計算社会科学

キーワード：オルトメトリクス　ソーシャルメディア　計量書誌学　研究評価　学術雑誌　プライバシー　固有表現抽出　学術情報流通

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９、ＣＫ－１９（共通）

１．研究開始当初の背景

50 年以上も前から学術研究の定量評価が試みられており，この評価は主に，学術文献の被引用数をもとにしている。しかし，学術文献が引用され始めるまでに数年の時間が必要であり，ある研究が社会的に注目されていたとしても，その注目を定量評価できるのは，ずいぶんと後のことである。2010 年前後より，被引用数にかわる指標としてオルトメトリクス（Altmetrics）が注目されており，文献の閲覧数，ブログやソーシャルメディアでの言及，マスメディアでの報道など，社会的な影響を加味した文献評価指標である。この指標には，引用文献の出版を待たず，早期に影響度を算出できるという利点があるとされているものの，現状では，言及数を単純にカウントしただけであり，言及元媒体が有するサービス固有の特徴的データを十分に生かしていない。

ソーシャルメディアで言及される学術文献はどのような文献であるか，という調査研究も行われている。このような調査研究は，社会的な影響力の大きい学術文献の特性を明らかにしようとするものであるが,従来調査の大半は PubMed を中心とする医療に関する文献の分析に偏り,そのような特定の分野に関する分析においても言語が英語に限定されている。オルトメトリクスはインパクトファクターと異なり,分野を横断しての評価が可能であるとされているものの,分野を横断した定量的な調査が行われておらず，その可能性が適切に検証されていない。

２．研究の目的

本研究では，新たな学術文献評価指標を開発することを目指し，ソーシャルメディアにおける学術文献の言及要因を検証した上で，その言及量の予測モデルを構築する。研究代表者のこれまでの研究により，ソーシャルメディアは世間の流行をリアルタイムに捉えていることが明らかになっている。また，予備調査により，ソーシャルメディア上での学術文献に関する言及が増加していることも確認した。これらより，学術文献評価にもソーシャルメディアのデータを有効に活用できると考えた。

まず，PubMed を中心とした医療に関する文献に限定せず，分野を横断した言及要因調査を行う。この目的を達成するために，言及情報及び文献情報を大規模に収集する。従来研究ではデータ収集の都合により限られた一部のデータのみを分析対象としていたが，一部のサンプリングデータのみならず概ね全数のデータを利用した網羅的分析が可能である。さらに，情報の拡散のしやすさに基づいた言及量予測モデルの構築を行う。このモデルでは，あるコミュニティでのみ言及されたのか，それとも様々なコミュニティで言及されたのかを定量的に区別できることを目指す。

３．研究の方法

本研究では，主に(1)データ収集範囲の拡張，(2)言及要因の調査，(3)情報伝播経路をもとにしたオルトメトリクスの算出，の 3 項目に取り組む。各項目の成果を他の項目にも反映するため，並行して実施する。

(1)データ収集範囲の拡張

学術文献の提供データベースや言語を問わずに，網羅的に文献言及情報を収集するシステムを構築する。外国語文献には DOI（Digital Object Identifier）が付与されているケースが大半であり，DOI リンクをクリックした先のホスト名（URL の一部）を調査した後，そのホスト名を各言及源（たとえば Twitter）の検索システムで検索することにより，言及情報を収集する。

我が国の大学現場ではソーシャルメディアなどの個人による情報発信の盛り上がりよりも，ニュース報道など準公的な露出が重視されるケースが多いため，我が国の実態に合った学術評価のためにはニュース記事からの言及も適切に捉える必要がある。ニュース記事では通常，学術文献へのリンクが張られないため，記事の内容をもとに該当文献との関連づけを行う必要がある。学術雑誌名は明記される傾向にあるため，学術雑誌名及び報道日をもとに，記事と文献との関連付けアルゴリズムの開発を行う。

(2)言及要因の調査

どのような文献が言及されやすいかを調査する。研究計画時点では，以下の観点からの調査を予定している。これらの調査により，文献情報そのもの情報からのみで言及量を予測可能であるかを検証する。

・オープンアクセス論文は言及されやすいか
・言及されやすい分野はどのような分野か
・言及源ごとに言及されやすい分野が異なるか
・言及と被引用数に関係を見いだせるか
・文献タイトルから言及数の予測を立てられるか

(3)情報伝播経路をもとにしたオルトメトリクスの算出

ソーシャルグラフ上に言及ユーザをマーキングし，そのマーカーがどのように分布しているかに着目する。その分布をはかる指標として，Closeness Centrality を拡張し，マーカー間の最短距離の平均を使う方法を検討する。この値は，広く伝搬していれば値は大きくなり，特定のユーザの周囲にしか伝搬していなければ値が小さくなることが見込まれる。その結果，多数のコミュニティで言及されたか否かの判別が可能になり，高度な言及量の予測モデルを構築できると考える。

４．研究成果

(1)データ収集範囲の拡張

データの収集範囲を拡張し，日本の学術文献について，文献データベースとソーシャルメディアサービスのそれぞれにおいて，10 以上のサイトに対応させた。これらの成果については，Ceek.jp Altmetrics（http://altmetrics.ceek.jp/）で閲覧可能であり，リアルタイムに学術文献に関する社会の言及状況を把握できる（図 1）。日本の文献に関するソーシャルメディアの言及を収集し，分析，閲覧可能なウェブサイトを提供しているのは研究代表者らによるサイトのみである。海外のデータに関しては，arXiv.org，ACM Digital Library，IEEE Xplore Digital Library の収集に対応したため，今後，リアルタイムな表示にも対応していく。

ニュース記事と文献との関連付けにおける雑誌名の抽出について，国立国会図書館や学術文献データベースに登録されている雑誌名のリスト（辞書）との単純照合で済むと予想していたものの，実データで予備実験を実施したところ，特に外国語雑誌において表記揺れが多いことが明らかになり，雑誌名のリストの拡張および自動抽出を検討した。自然言語処理におけるブートストラップ法を用いる際に，尤度比を保守的に見積もることにより，高精度に抽出するアルゴリズムを実現した。これらの成果は，主な発表論文等の学術雑誌①②などで公表済みである。今後，記事と文献との関連付けに注力していく。

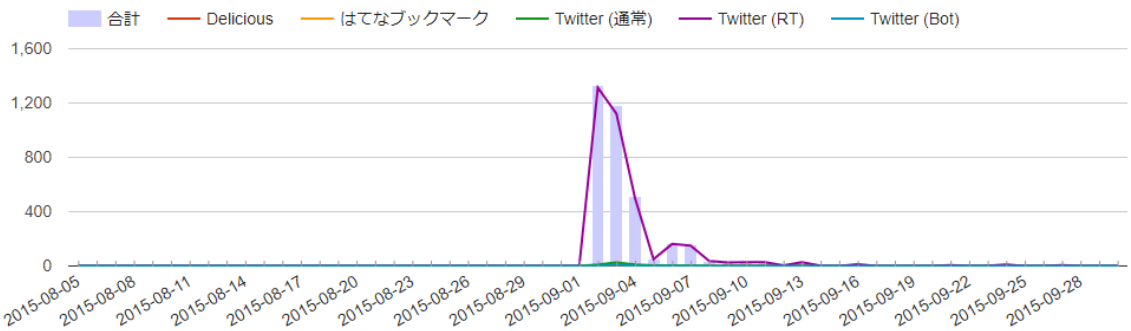

図 1：言及量グラフ表示の例

(2)言及要因の調査

日本の学術文献を対象とした言及要因の調査を行い，次に述べる知見などを得た。まず，ソーシャルメディアで言及されている日本の学術文献は 1.0%～2.2%程度であり，日本国外を対象とする従来の調査結果（10%～20%程度の文献が言及される）よりも小さな値であった。また，様々なソーシャルメディアのうち，Twitter からの言及量がほかよりも著しく多かった。そして，従来の調査結果と異なり，本研究で作成したデータセットにおいては，ソーシャルメディアでの言及量と論文出版年との関係は見いだせなかった（表 1）。最後に，理工学系の論文よりも人文社会学系の論文の方が，ソーシャルメディアでよく言及されることが明らかになった。これらの成果は，主な発表論文等の学術雑誌⑥⑧⑫などで公表済みである。この調査では，従来の調査結果の一部を覆しており，今後，海外の文献にも調査範囲を広げ，より詳細に分析していく。

表 1：CiNii Articles から論文を抽出し，ソーシャルメディアで言及されているかを調査

出版年	被言及論文数	被言及論文の割合	全論文数
2006	1,293	1.04%	124,108
2007	1,337	1.11%	120,585
2008	1,439	1.16%	123,883
2009	1,411	1.17%	120,964
2010	1,258	1.08%	116,758
2011	1,084	1.08%	100,746
2012	1,105	0.99%	111,433
2013	949	0.95%	100,022
2014	828	0.97%	85,665
2015	341	0.44%	76,676

(3)情報伝播経路をもとにしたオルトメトリクスの算出

ソーシャルグラフ上で情報拡散が生じるという仮説を立て，カスケードモデルによる情報拡散モデルを構築し，言及量の予測を試みた。しかしながら，収束した言及量の予測については，ソーシャルグラフにおけるカスケードモデルよりも，初期言及量の方がモデルのあてはまりがよかった。そのため，ソーシャルグラフ上で情報拡散が生じるという仮説を積極的に支持する結果が得られていない。

情報伝搬に関する新たな仮説を立てて検証を続けるよりも，ソーシャルメディア利用の状況を詳細に分析する方が今後に繋がると考え，情報伝搬経路をもとにしたオルトメトリクスの算出を断念し，ソーシャルメディア利用ユーザの分析を行なうこととした。

居住地推定を題材とし，場所に関する情報を投稿せずとも，一見場所とは関係のない天気に関する情報でも居住地の推定が可能であることが分かり，ユーザのプライバシーが意図せず漏れている可能性があることが示唆された（学術雑誌⑩）。さらに，このようなプライバシーがプロフィールなどに記載されるのかどうかを日英の言語別に調べたところ，日本語ではユーザの現在の状況が，英語ではユーザの誕生日などが記載される傾向にあることが明らかになった(学術雑誌⑤)。このほかにも，引っ越しなどの環境の変化におけるソーシャルグラフの収束時期を調査したところ，事象が発生してから半年から１年後にソーシャルグラフの変化が収束することが分かった（学術雑誌④）。また，ソーシャルグラフ構築の際のデータ活用方法の検証（学術雑誌⑨），政治情報の情報伝搬の分析（学術雑誌③）も行っている。今後，以上で得られた知見をもとに，オルトメトリクスの適切な算出方法を模索する。

５．主な発表論文等

〔雑誌論文〕（計 12 件）

① 菊地真人，川上賢十，吉田光男，梅村恭司．観測頻度に基づくゆう度比の保守的な直接推定．電子情報通信学会和文論文誌 D．vol.J102-D，no.4，pp.289-301，2019．
https://doi.org/10.14923/transinfj.2018DEP0007 （査読あり）

② Masato Kikuchi, Mitsuo Yoshida, Kyoji Umemura. Journal Name Extraction from Japanese Scientific News Articles. Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2018. pp.143-148, 2018.
https://doi.org/10.23919/APSIPA.2018.8659765 （査読あり）

③ Mitsuo Yoshida, Fujio Toriumi. Information Diffusion Power of Political Party Twitter Accounts During Japan's 2017 Election. Proceedings of the 10th International Conference on Social Informatics. vol.2, pp.334-342, 2018.
https://doi.org/10.1007/978-3-030-01159-8_32 （査読あり）

④ Shiori Hironaka, Mitsuo Yoshida, Kyoji Umemura. Temporal Analysis of Online Social Graph by Home Location. Proceedings of the ACM IUI 2018 Workshop on Web Intelligence and Interaction. 2018.
http://ceur-ws.org/Vol-2068/wii11.pdf （査読あり）

⑤ Jinsei Shima, Mitsuo Yoshida, Kyoji Umemura. When Do Users Change Their Profile Information on Twitter?. Proceedings of the 2017 IEEE International Conference on Big Data. pp.3119-3122, 2017.
https://doi.org/10.1109/BigData.2017.8258287 （査読あり）

⑥ 吉田光男．集合知による新たな研究評価．Biophilia．vol.6，no.3，pp.31-37，2017．
http://biophilia.jp/journal/biophilia-23-1.html

⑦ Yuki Kondo, Masatsugu Hangyo, Mitsuo Yoshida, Kyoji Umemura. Home Location Estimation Using Weather Observation Data. Proceedings of the 2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications. 2017.
https://doi.org/10.1109/ICAICTA.2017.8090972 （査読あり）

⑧ 佐藤翔，吉田光男．日本の学協会誌掲載論文のオルトメトリクス付与状況．情報知識学会誌．vol.27，no.1，pp.23-42，2017．
https://doi.org/10.2964/jsik_2017_009 （査読あり）

⑨ 廣中詩織，吉田光男，岡部正幸，梅村恭司．日本における居住地推定に利用するためのフォロー関係の調査．人工知能学会論文誌．vol.32，no.1，pp.WII-M_1-11，2017．
https://doi.org/10.1527/tjsai.WII-M （査読あり）

⑩ Yuka Kamiko, Mitsuo Yoshida, Hirotada Ohashi, Fujio Toriumi. Uncovering Information Flow Among Users by Time-Series Retweet Data: who is a friend of whom on Twitter?. Proceedings of the 2016 IEEE International Conference on Big Data. pp.2500-2504, 2016.
https://doi.org/10.1109/BigData.2016.7840888 （査読あり）

⑪ Shiori Hironaka, Mitsuo Yoshida, Kyoji Umemura. Analysis of Home Location Estimation with Iteration on Twitter Following Relationship. Proceedings of the 2016 International Conference On Advanced Informatics: Concepts, Theory And Application. 2016.
https://doi.org/10.1109/ICAICTA.2016.7803100 （査読あり）

⑫ 吉田光男．ソーシャル言及数で論文に新たな評価軸．日経ビッグデータ．2016 年 6 月号（no.28），pp.29，2016．
http://business.nikkeibp.co.jp/atclbdt/15/258685/052300013/

〔学会発表〕（計 6 件）
① 小林和央，風間一洋，吉田光男，大向一輝．インターネット上の論文の閲覧行動と言及行動の関係の分析．第 3 回計算社会科学ワークショップ．2019．
② 吉田光男．ウェブマイニングのためのデータ収集と学術情報分析．第七回計算社会科学とその周辺セミナー（招待講演）．2019．
③ 吉田光男．共著者ネットワークをもとにした共同研究者探索支援システムの試作．第 12 回 Web インテリジェンスとインタラクション研究会．2018．
④ 鳥海不二夫，吉田光男．2017 年衆議院選挙における政党公式アカウントフォロワーの分析．第 32 回人工知能学会全国大会．2018．
⑤ 島仁誠，吉田光男，梅村恭司．Twitter のプロフィール変更時におけるユーザの行動と関係のあるキーワードの抽出．第 11 回 Web インテリジェンスとインタラクション研究会．2017．
⑥ 廣中詩織，吉田光男，岡部正幸，梅村恭司．日本における居住地推定のためのソーシャルネットワーク作成方法の調査．第 30 回人工知能学会全国大会．2016．

〔図書〕（計 0 件）

〔産業財産権〕
○出願状況（計 0 件）

○取得状況（計 0 件）

〔その他〕
ホームページ等
http://altmetrics.ceek.jp/

６．研究組織