オルトメトリクスを用いた尖った研究発掘システムの開発

松本　馨（九州大学　インスティテューショナル・リサーチ室）

１.はじめに

九州大学では 2018 年度に研究力データ等に基づいた分析結果を参考にした研究フラッグシップ（研究上の強み）を策定し，指定国立大学法人として強い分野の研究力を維持，向上していくことを目標に掲げている．さらに，その他の強みを生み出していくことも必要と考えており，研究の多様性を確保して将来への投資をする観点から，これまでになかった埋もれていた研究（尖った研究）を発掘し，組織としてそれを支援して新たな研究上の強みにして，社会実装に繋がる成果を創出することを目指している．

2019 年度には，インスティテューショナル・リサーチ室で尖った研究の発掘分析が企画・実施され，その成果としてインターネットを情報源とした新しい指標であるオルトメトリクス（Altmetrics）を利用した研究発掘の仕組みを提案した．現在は，それを用いたシステムの開発に着手し，研究発掘で活用されることを目指している．

２.オルトメトリクス

２.１　定義と特長

オルトメトリクスとは Alternative と Metrics を組み合わせた造語で，ソーシャルメディアを活用して研究成果のインパクトを論文レベルでリアルタイムに計量化する新たな研究評価指標である[1]．広義にはインターネットを利用した仕組み全般が含まれており，従来手法である論文の被引用数（Citations）や，論文ダウンロード数等の閲覧統計，論文管理ツールの登録数を用いるものもある．また，どのサービスでも Twitter や Facebook 等のソーシャルメディアを集計対象にしている[2]．これらは，インターネット上のサーバーで自動集計されて指標の反応・反映が早いことから，従来手法である Citations に反映されるよりも前に，その論文が社会にもたらした影響を捉えることが出来るのも利点である．

２.２　代表的なオルトメトリクス・ツール

現在，主に利用されているオルトメトリクス・ツールには，次の 2 社のものがある[3]．

（a）Altmetric

英国の出版社 Macmillan の一部門 Digital Science のグループ企業である．DOI や PMID 等研究成果を特定する識別子に基づき，ソーシャルメディア，ブログそしてウィキペディア等 Web 上での言及等を定量的に測定し，Altmetric Attention Score を提供している．

（b）Plum Analytics

機関向けに PlumX Print を提供する製品 PlumX Metrics を開発している．2014 年，EBSCO の完全子会社となり，ディスカバリーサービス EBSCO Discovery Service や論文データベース EBSCO host に PlumX Print が導入された．2017 年，Elsevier に買収され，現在は同社によってサービスが提供されている．

２．３　PlumX Metrics

九州大学では，既に Elsevier が提供するサービス（Scopus, SciVal, Pure）を契約していることから，そのまま利用出来る PlumX Metrics を試用した．PlumX Metrics は，表 1 のように Social Media，Mentions，Captures，Usage，Citations の 5 指標で構成されている．

表 1：PlumX Metrics の構成

指標名	指標の概要
Social Media	ソーシャルメディア（主に Twitter, Facebook）での言及数
Mentions	広報メディア等による紹介記事数
Captures	論文管理ツール（主に Mendeley）のブックマーク数
Usage	論文ファイル（主に EBSCO, PLoS）の閲覧・ダウンロード数
Citations	論文の被引用数（従来手法）

２．４　指標の特徴調査

PlumX を試用する前に，各指標の統計値を調査した．基本統計量が表 2，各指標間の相関係数が表 3 である．この結果から，指標値の分布幅が最小で 0～436，最大で 0～108062 もの差があり，これらの指標を同じように使うのは難しいことが分かった．さらに，Citations と Captures の相関係数が 0.659 とやや強い相関を示している一方で，その他の指標とは 0.072～0.098 であり，ほとんど Citations と相関がないことも分かった．

表 2：PlumX 各指標の基本統計量

	度数	最小値	中央値	平均値	最大値	標準偏差
Social Media	28307	0	0	2.54	5499	57.98
Mentions	28307	0	0	0.13	436	3.50
Captures	28307	0	3	19.54	6894	89.40
Usage	28307	0	1	142.12	108062	1433.53
Citations	28307	0	5	23.66	4635	78.85

表 3：PlumX 各指標間の相関係数

		Social Media	Mentions	Captures	Usage	Citations
Social Media	Pearson の相関係数	1	.505**	.224**	.193**	.075**
	有意確率（両側）		.000	.000	.000	.000
	N	28307	28307	28307	28307	28307
Mentions	Pearson の相関係数	.505**	1	.184**	.095**	.098**
	有意確率（両側）	.000		.000	.000	.000
	N	28307	28307	28307	28307	28307
Captures	Pearson の相関係数	.224**	.184**	1	.143**	.659**
	有意確率（両側）	.000	.000		.000	.000
	N	28307	28307	28307	28307	28307
Usage	Pearson の相関係数	.193**	.095**	.143**	1	.072**
	有意確率（両側）	.000	.000	.000		.000
	N	28307	28307	28307	28307	28307
Citations	Pearson の相関係数	.075**	.098**	.659**	.072**	1
	有意確率（両側）	.000	.000	.000	.000	
	N	28307	28307	28307	28307	28307

**. 相関係数は 1% 水準で有意（両側）です。

Citations と Captures の相関が強いことは注目に値する結果である．Captures のデータソースには Elsevier が提供する論文管理ツール Mendeley のブックマーク数があり，これが Citations との強い相関を示す要因になっていると考えられる．同じようなことは文献[4]で PDF/html ダウンロード数との相関が 0.34～0.65，Mendeley 及び CiteULike でのブックマーク数との相関が 0.15～0.43 出ているとの結果が示されており，九州大学の論文についてもこれらと同様の傾向が出ていると考えられる．

その他の指標について例を挙げると，文献[5]では Twitter の Tweet 数と Citations との関係を調査しており，その結果によれば論文公表直後 3 日間の Tweet 数が多い論文は，後で Citations が上昇する傾向にあることが示されている（図 1）．

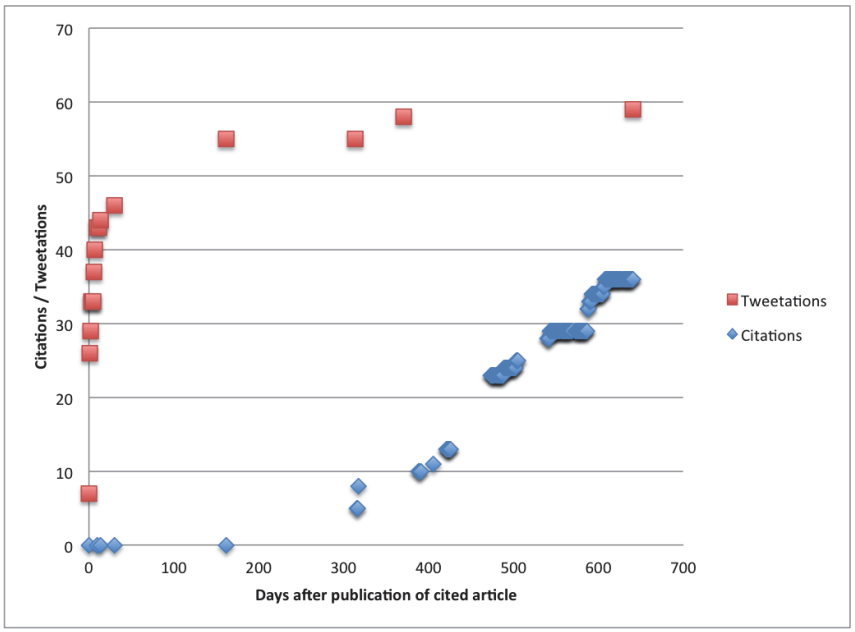


図 1：Citation and tweetation dynamics of a highly cited (and highly tweeted) article [article ID 1376]; citations according to Scopus.(文献[5])

２．５　指標の活用方法検討

以上から，Citations とやや強い相関を示す Captures を Citations の先行指標として用い，その他の指標は社会的な影響を多様な観点から捉えるための指標として扱い，様々な観点から尖った研究を発掘することを考えた．

しかし，インターネットからの情報では個人を特定することが困難で，本来の研究とは無関係の素人によるアクセスが相当数，含まれることが懸念される．具体的には，各指標の特性から，次のような事が考えられる．

- Social Media はアクセスが容易で，素人の反応が多く含まれる
- Mentions は素人は少ないが，広報メディアによる PR を意図した記事が含まれる
- Captures は，ユーザ登録をして専用ツールをインストールする必要があるため専門家が多い
- Usage は，オープンアクセスの論文に対しては一般的な Web ブラウザで見られるためアクセスが容易で，素人が含まれる

さらに文献[6]で指摘されている通り，オルトメトリクスが利用するプラットフォームはそれぞれ目的と機能が異なるため生成される指標の意味が異なることや，ソーシャルメディアに基づく指標は一般受けしやすい記事に反応が偏る傾向や，指標そのものの自己操作が容易であることも問題である．これらの事情を考慮して，最終結果は各部局の専門家に精査してもらうことが必須と考えた．

そこで，専門家の負担を軽減するために素人を極力除外するフィルタ設定を行い，さらに，各指標を使って十分に絞り込んだ結果リストを提示して，専門家による判断を仰ぐことを考えた．フィルタには下記の指標を用いることにして，最終的には図 2 の手順で尖った研究発掘を試行した．

- 著者数：著者数が多い論文は既にメジャーと考えられることから，著者数の少ない論文を対象にする
- 出版年：今後伸びると思われる新しい論文を探すために，古い論文は対象外にする
- FWCI（Field Weighted Citation Impact）：被引用数の大きな論文は既にメジャーと考えられることから，被引用数の小さい論文を対象にする
- FWVI（Field Weighted Views Impact）：論文の表示回数（ダウンロード数）が大きいものは，何らかの理由で注目されている論文である

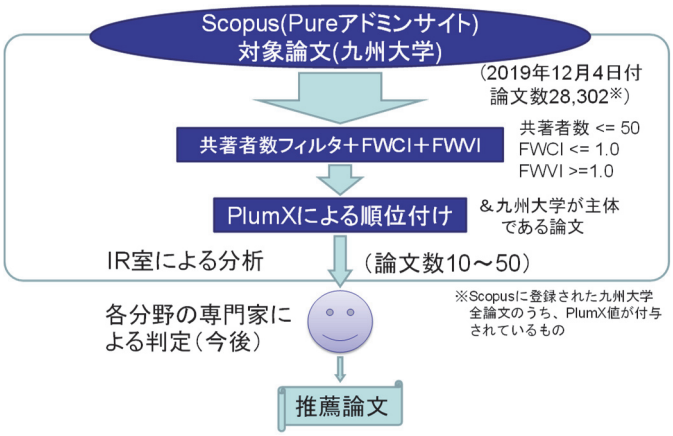


図 2：尖った研究発掘の流れ

3．システム化

試行の結果，オルトメトリクスを学内に認知させることが出来た．さらに，実際に利用して次の点が課題として考えられた．

- 値の変化が急激で，毎月，多種多様な論文が逐次公開されることから，一度，報告するだけでは意味を成さないこと
- 単独で値を見ても，その大小を判断することが難しいこと（実績の無い指標であり，どの程度の値を目指すべきか，参考値も存在しない）
- 文献タイプによって値の出方が大きく変わること
- 無料で閲覧できるオープンアクセス論文との区別をしていなかったこと

これらへ対応するために，定期的に指標データの取り込みを行って値の変化を追うことが出来るようにして，時系列変化を追跡することで，どの論文が尖っているか見分けられ

るようにすることが必要と考えた．そこで BI ツール（Tableau）による指標の可視化を行って，そのダッシュボードを学内の各部局と共有し，専門家がいつでも指標データを閲覧できる尖った研究発掘システムの開発を行った．

システムで用いるデータは，図 3 の流れで処理される．まず，IR 室スタッフ作業によるデータ取得を定期的に行い，指標データを Excel 形式で保管する．続いて，蓄積したデータを Tableau に流し込んで最新値を追加する．最新になった Tableau データを Tableau Server へアップロードする．アップロードし公開されたデータを各部局で参照し，尖った研究発掘を行う．尖った研究発掘システムの画面例を図 4 に示す．データ取得日，文献タイプ，オープンアクセスの有無，著者数，出版年，FWVI，FWCI，ASJC 分類，部局名でフィルタをかけることが出来るようになっている．

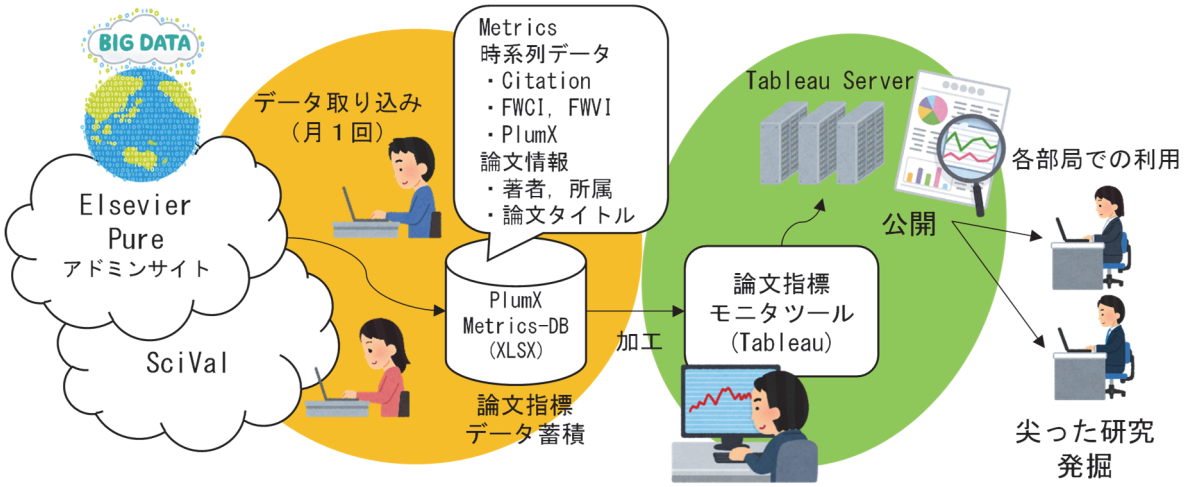


図 3：尖った研究発掘システム指標公開までの流れ

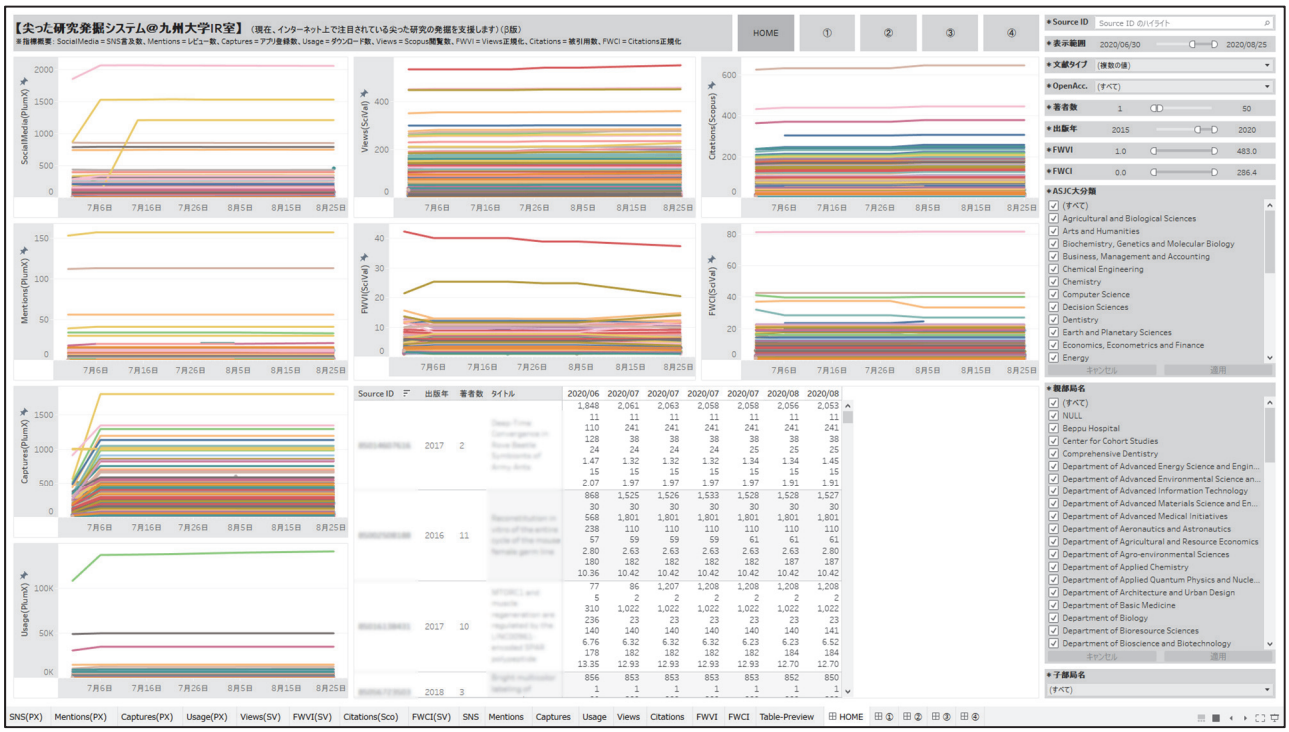


図 4：尖った研究発掘システム画面例

４.利用事例と今後の課題

利用事例としては，総務部広報室が研究成果を英文プレスリリースプラットフォームに投稿し，掲載された記事のインパクトを定量的に測る方法として PlumX 値を用いた例が挙げられる．尖った研究発掘システムは開発後，間もない状況であり，積極的な利用が期待される．今後の課題には次の事項が挙げられる．

① 本システムの対象論文は DOI が設定され，Scopus, Pure に登録されている論文に限られること（実質的に英文誌に限られ，人文社会科学系の大半に対応しない）
② 結果の妥当性が不明であること（特に Citations と Captures には相関関係はあるが，因果関係があるか不明）
③ フィルタの妥当性が不明であること
④ データ更新が 1 ヶ月毎の手動取得であること

このうち，①の人文社会科学系への対応では和文対象のオルトメトリクスを用いることが考えられる．和文を対象にしたオルトメトリクスには Ceek.jp Altmetrics[7]があり，九州大学関連の論文は 2,182 件とまだ多くないが，登録数が増えて他のデータベースとの突合が出来るようになれば．活用可能性は大きく広がるものと考えている．②については，専門家による結果の確認を行うことで最終結果の質を担保し，活動を通じて妥当性検証と有効な指標の特定が進むことを期待したい．③については，②と並行してトライ&エラーの繰り返しによるフィルタの見直しを図っていく必要がある．④については，より高いニーズが生まれてくるようであれば，Web API を利用した自動化が必要になると考えている．

本システムはデータの蓄積が始まったばかりである．蓄積が進むことで過去の指標値とその後の結果を照らし合わせることが可能になり，ここで検討した仮説の妥当性検証や，各指標の性質が明らかになる．それにより積極的な指標活用が進んで研究発掘が行われ，最終的には，社会実装に繋がる研究テーマが開拓されることを期待している．

【参考文献】

[1] 坂東慶太(2012)，“Altmetrics の可能性 ソーシャルメディアを活用した研究評価指標”，情報管理，Vol.55，No.9，pp.638～646
[2] 佐藤翔，吉田光男(2016)，“オルトメトリクスは論文評価を変えるかーソーシャルメディアで算出する新たな指標”，化学，Vol.71，No.2，pp.23～28
[3] 坂東慶太(2017)，“ごぞんじですか？ 第 104 回 altmetrics の概要と近年の動向について”，専門図書館，No.283，pp.40～43
[4] Jason Priem, Heather A.Piwowar, Bradley M.Hemminger(2012),“Altmetrics in the Wild: Using Social Media to Explore Scholarly Impact”, arXiv(Computer Science > Digital Libraries), https://arxiv.org/abs/1203.4745v1
[5] Gunther Eysenbach(2011),“Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientifc Impact”, Journal of Medical Internet Research, Vol.13, No.4, e123, DOI: 10.2196/jmir.2012
[6] 孫媛(2017)，“研究評価のための指標：その現状と展望”，情報の科学と技術，67 巻，4 号，pp.179～184
[7] 吉田光男(2014)，“計量書誌学の新たな挑戦－国産オルトメトリクス計測サービスの開発－”，情報の科学と技術，64 巻，12 号，pp.501～507