

令和 5 年 6 月 21 日現在

機関番号：12102

研究種目：基盤研究(B) (一般)

研究期間：2019～2022

課題番号：19H04421

研究課題名(和文) 利用者の研究練度に応じた多様な観点を統合する学術情報システム

研究課題名(英文) Academic information system that integrates various viewpoints according to users' research skill

研究代表者

吉田 光男 (Yoshida, Mitsuo)

筑波大学・ビジネスサイエンス系・准教授

研究者番号：60734978

交付決定額(研究期間全体)：(直接経費) 12,300,000円

研究成果の概要(和文)：本研究では、利用者の状況に応じ、多様な観点で学術情報を提示できる学術情報システムを実現することを目指し、利用者の様々な探索要求に対応する新しい学術情報評価指標を複数開発した上で、利用者の行動履歴をもとに研究練度を推定し、利用者の研究練度に応じて複数の指標を自動的に統合する学術情報システムの構築に取り組んだ。その結果、論文に対する言及パターンから論文の普遍度を測る指標を開発できた。また、言及パターンとアクセスパターンを特徴量とする表示アルゴリズムを開発し、有効性を検証できた。

研究成果の学術的意義や社会的意義

現在の学術情報システムは研究者の高度な専門性を前提とした画一的なシステムであり、それ以外の利用者の要求に応えられていないところ、多様な学術情報評価指標を開発したことにより、利用者の利用状況に応じた学術情報の提示が可能となる。また、本研究は学術情報システムを対象とするものの、検索対象文書、対象文書に対する閲覧・言及などの動作、利用者の行動履歴とを組み合わせるものとして一般化することができ、学術情報以外への応用も可能である。

研究成果の概要(英文)：In this study, we aimed to realize an academic information system that can present academic information from various viewpoints according to the user's situation. We developed several new indicators for evaluating academic information that correspond to users' various search requests, and then worked on the construction of an academic information system that estimates the level of research skill based on users' behavioral history and automatically integrates several indicators according to users' level of research skill. As a result, we developed an index that measures the universality of an article based on its mention pattern. In addition, we developed a display algorithm that uses mention patterns and access patterns as features, and verified the effectiveness of the algorithm.

研究分野：計算社会科学

キーワード：オルトメトリクス 学術情報 研究評価 情報検索 ソーシャルメディア Twitter CiNii Articles

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

学術情報(論文)の評価は、論文のメタデータ(論文の新しさなど)、論文の内容(ピア・レビューなど)、論文の引用関係(被引用数など)を用いて行なわれ、特に定量評価では、被引用数が用いられる。被引用数は、学術雑誌の評価(インパクト・ファクター)や研究者の評価(h-index)など、様々な評価の元となっている。しかしながら、被引用数には、研究分野や主題によって論文の引用のされやすさが異なり、さらに、研究者による評価観点しか考慮されていないという問題を抱える。近年では、これらを補完する指標として、学術情報の閲覧数、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オルトメトリクス(Altmetrics)が注目されている。我々は、日本で初めてオルトメトリクスの大規模な調査を行ない、被引用数と異なる傾向があることを明らかにした。この調査では、オルトメトリクスによる学術情報の評価について、一部の専門家の視点ではなく、学術情報を必要とするより多くの利用者の視点を捉えられることを示唆している。

論文出版数をもとに研究経験を5段階(Unpublished, Postgraduate, Postdoc, Lecturer, Professor)に分け、それぞれ独立な4種類の論文推薦システムを適用したところ、研究経験段階によって異なる推薦方法が適していることが明らかになっている。また、文献利用者を12属性(Student, Lecturer, Professor, Librarianなど)に分け、保存している文献の種類を調査したところ、属性ごとに異なる傾向があることも明らかになっている。これらつまり、利用者の研究経験の度合いによって、提示すべき論文が異なることを示している。このことは、研究者が専門分野以外を調べる場合(経験は十分だが、対象分野の知識がない場合)、開発者が専門分野の論文を調べる場合(対象分野の知識があるが、研究経験が浅い場合)なども同様である。つまり、論文を画一的に扱うだけでは不十分で、利用者の対象分野に対する知識や研究経験の度合い(まとめて「研究練度」と呼ぶ)に応じて、異なる扱いをしなければいけない必要性が示されている。しかしながら、現在の学術情報システムでは、表示順序の初期値として文献の出版日(新着順)が用いられることが多く、どのような利用者であっても画一的な表示順序が採用されている。

2. 研究の目的

本研究の目的は、利用者の状況に応じ、多様な観点で学術情報を提示できる学術情報システムを実現することである。この実現のために、様々な利用者の探索要求に応えられる新たな学術情報評価指標を複数開発した上で、利用者に応じて複数の指標を自動的に統合する学術情報システムを構築する。さらに、既存の実運用サービスと連携し、様々な利用者に対して本研究の有用性を検証する。

3. 研究の方法

データ基盤の準備

本研究を遂行するためには、学術情報データ(メタデータ・被引用数)、学術情報に対する言及データ、利用者の行動データ、の3種を必要とする。については国立情報学研究所のCiNiiのデータを利用し、についてはこれまでの研究で構築した収集システム(altmetrics.ceek.jp)を活用する。さらに、については科学報道記事における言及の収集に対応させ、また、についても詳細な行動データを取得する仕組みを構築する。

学術情報に関する多面的評価指標

新たな評価指標の開発にあたり、まずは、これまでの調査の延長として、従来から良く用いられている被引用数と、閲覧・言及・報道それぞれの回数との関係を明確にする。被引用数は研究者の専門家、報道回数は報道の専門家による注目度である一方、閲覧および言及は、より広範囲な利用者による注目度である。また、閲覧内容は他人に知られることはないが、言及内容は他人に知られる(あるいは知らせる意図がある)という違いがある。それぞれの指標間での相関を調査した上で、指標の高低と論文内容との関係を分析し、指標開発の基礎とする。

多面的評価を扱える学術情報システム

現在の文献検索システムは、一般的に、検索キーワードを含む文献集合を抽出し、出版日や被引用数などでソートして文献を提示する。文献集合を抽出とソートはそれぞれ独立した処理であり、ソートも単一指標に基づく画一的なソートである。本研究ではまず、利用者の研究練度を推定する手法を構築し、その推定練度に応じてソートを最適化するアルゴリズムを追求する。そして、このアルゴリズムの追求が一段落したらば、検索キーワードとの適合性も考慮したソートを検討する。

4. 研究成果

本研究によって得られた主要な成果は、以下の通りである。

データ基盤の準備

arXiv.org や bioRxiv.org などのプレプリント、ACM Digital Library や IEEE Xplore などの海外の電子ジャーナル論文に対する、ソーシャルメディアでの言及を収集するなど、研究対象とするデータの範囲を広げた。また、ニュース収集システムを改良し、最新の文字コードに対応するようにした。

学術情報に関する多面的評価指標

非専門家にとっての論文タイトルの「面白さ」を得点化し、ソーシャルメディア（Twitter）からの言及数が多い論文と言及されたことのない論文で、この得点に差があるかを分析した。その結果、Twitter からの言及数が多い論文グループと、言及されたことのない論文グループとで「面白さ」得点には有意な差が存在し、言及数が多い論文の方がタイトルが「面白い」傾向が確認された。

arXiv 論文に対する Twitter での言及特性を分析した。その結果、arXiv 論文の情報流通において、ボットが重要な役割を果たしていることが明らかになり、さらに、情報拡散者の関係ネットワークから、近年急速に発展している機械学習分野について、積極的な貢献をしている国際グループと日本グループの存在が確認された。

論文がどの程度普遍的なものとして扱われているかを示す普遍度を開発した。この指標は、論文に対する言及や閲覧のコミュニティ構造をモデル化したものであり、CiNii Articles の検索ログとソーシャルメディア上の論文の言及データを用い、同一の論文集合に対する論文の閲覧と言及の違いや特性を分析すると同時に、提案指標の妥当性を検証した。その結果、普遍度が高い場合は、ユーザ数も多い傾向があるものの、逆は成り立たず、ソーシャルメディア上で話題性だけで広く拡散するだけの論文の評価を下げられることなどが分かった。また、普遍度はユーザ集合の年齢や職業などの影響を受けていることも分かり、多面的な評価観点が取り入れられていることが示唆された。

多面的評価を扱える学術情報システム

多面的評価を扱える、つまり複数のソートを導入できる検索システムの基盤を構築した。具体的には、書誌情報がオープンデータとして公開されている arXiv.org を対象とし、本研究課題で開発している評価指標の一部で並び替えることのできる論文検索システムを構築した。また、arXiv.org の言及者を 4 つの行動タイプに分け、言及の特徴を分析したところ、それぞれのユーザは、4 タイプのいずれかの行動タイプのユーザとして振る舞い、複数の行動を行わない傾向にあることが分かった。

オルトメトリクスで用いられる論文評価指標の時系列ベクトル表現から検知したバーストの有無や期間・回数などのバースト特性から、多様なユーザによって生み出された論文の評価について多面的に分析した。具体的には、2 年間のソーシャルメディアの論文言及データから論文言及人数、CiNii Articles の検索ログから論文閲覧人数を求め、それらの時系列変化から言及バーストと閲覧バーストを検知し、論文とそのバースト特性の関係を分析した。その結果、閲覧人数では特定の専門分野の研究者や学生向けよりも、ソーシャルメディア上の情報拡散の影響を強く受けた一般的なユーザ向けの論文がより上位にくることと、ソーシャルメディアの情報拡散と学術イベントが引き起こすバースト特性は異なることが分かった。

5. 主な発表論文等

〔雑誌論文〕 計14件（うち査読付論文 9件 / うち国際共著 0件 / うちオープンアクセス 7件）

1. 著者名 小林和央, 風間一洋, 吉田光男, 大向一輝, 佐藤翔, 桂井麻里衣	4. 巻 J105-D(5)
2. 論文標題 論文のオルトメトリクスに関するバースト特性分析	5. 発行年 2022年
3. 雑誌名 電子情報通信学会和文論文誌D	6. 最初と最後の頁 310-321
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2021DEP0007	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 佐藤翔, 設楽成実, 矢野正隆, 北村由美	4. 巻 32(1)
2. 論文標題 『東南アジア 歴史と文化』誌掲載論文の引用文献の傾向と著者所属機関における所蔵状況	5. 発行年 2022年
3. 雑誌名 情報知識学会誌	6. 最初と最後の頁 73-87
掲載論文のDOI (デジタルオブジェクト識別子) 10.2964/jsik_2022_005	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Kiyosuke Shimada, Kazuhiro Kazama, Mitsuo Yoshida, Ikki Ohmukai, Sho Sato	4. 巻 -
2. 論文標題 Analysis of Leading Communities Contributing to arXiv Information Distribution on Twitter	5. 発行年 2021年
3. 雑誌名 Proceedings of the 20th IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology	6. 最初と最後の頁 40-47
掲載論文のDOI (デジタルオブジェクト識別子) 10.1145/3486622.3493947	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Tomoya Nishide, Marie Katsurai	4. 巻 -
2. 論文標題 Finding the Relevance Between Publication Venues Based on Research Trend Similarity and Citation Relationships	5. 発行年 2021年
3. 雑誌名 Proceedings of the 2021 ACM/IEEE Joint Conference on Digital Libraries	6. 最初と最後の頁 312-313
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/JCDL52503.2021.00047	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 小林和央, 風間一洋, 吉田光男, 大向一輝, 佐藤翔	4. 巻 30(3)
2. 論文標題 インターネット上のユーザの行動データを用いた論文の普遍性の分析手法	5. 発行年 2020年
3. 雑誌名 情報知識学会誌	6. 最初と最後の頁 312-327
掲載論文のDOI (デジタルオブジェクト識別子) 10.2964/jsik_2020_007	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 西岡千文, 佐藤翔	4. 巻 31(1)
2. 論文標題 Unpaywall を利用した日本におけるオープンアクセス状況の調査	5. 発行年 2021年
3. 雑誌名 情報知識学会誌	6. 最初と最後の頁 31-50
掲載論文のDOI (デジタルオブジェクト識別子) 10.2964/jsik_2021_016	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Takahashi Tetsuya, Tango Koya, Chikazawa Yuto, Katsurai Marie	4. 巻 -
2. 論文標題 A Novel Researcher Search System Based on Research Content Similarity and Geographic Information	5. 発行年 2020年
3. 雑誌名 Proceedings of the 22nd International Conference on Asia-Pacific Digital Libraries	6. 最初と最後の頁 390-398
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-64452-9_36	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 佐藤翔, 石橋柚香, 南谷涼香, 奥田麻友, 保志育世, 吉田光男	4. 巻 29(3)
2. 論文標題 Twitterからの言及数が多い論文は言及されたことのない論文と比べてタイトルが「面白い」	5. 発行年 2019年
3. 雑誌名 情報知識学会誌	6. 最初と最後の頁 268-283
掲載論文のDOI (デジタルオブジェクト識別子) 10.2964/jsik_2019_035	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 西岡千文, 亀田堯宙, 佐藤翔	4. 巻 30(1)
2. 論文標題 日本の学術出版物におけるオープン・サイテーションの分析	5. 発行年 2020年
3. 雑誌名 情報知識学会誌	6. 最初と最後の頁 3-20
掲載論文のDOI (デジタルオブジェクト識別子) 10.2964/jsik_2020_003	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

1. 著者名 吉田光男	4. 巻 38(3)
2. 論文標題 特集「研究評価と学術情報分析」にあたって	5. 発行年 2023年
3. 雑誌名 人工知能	6. 最初と最後の頁 358-359
掲載論文のDOI (デジタルオブジェクト識別子) 10.11517/jjsai.38.3_358	査読の有無 無
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

1. 著者名 佐藤翔	4. 巻 38(3)
2. 論文標題 査読におけるバイアスとその抑制の試み	5. 発行年 2023年
3. 雑誌名 人工知能	6. 最初と最後の頁 366-374
掲載論文のDOI (デジタルオブジェクト識別子) 10.11517/jjsai.38.3_366	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 桂井麻里衣	4. 巻 38(3)
2. 論文標題 学術データに基づく研究者の特徴表現	5. 発行年 2023年
3. 雑誌名 人工知能	6. 最初と最後の頁 392-398
掲載論文のDOI (デジタルオブジェクト識別子) 10.11517/jjsai.38.3_392	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 佐藤翔	4. 巻 73(1)
2. 論文標題 学術情報流通の多様化：査読誌の「4機能」の変化	5. 発行年 2023年
3. 雑誌名 情報の科学と技術	6. 最初と最後の頁 2-8
掲載論文のDOI (デジタルオブジェクト識別子) 10.18919/jkg.73.1_2	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 佐藤翔	4. 巻 67(1)
2. 論文標題 図書館総合展JPLA/JMLAフォーラム「医薬生物学分野におけるプレプリントをめぐる動向」	5. 発行年 2022年
3. 雑誌名 薬学図書館	6. 最初と最後の頁 7-12
掲載論文のDOI (デジタルオブジェクト識別子) 10.11291/jpla.67.1_7	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計15件 (うち招待講演 0件 / うち国際学会 0件)

1. 発表者名 吉田光男
2. 発表標題 COVID-19流行下でのプレプリントツイートに関する基礎調査
3. 学会等名 第35回人工知能学会全国大会
4. 発表年 2021年

1. 発表者名 吉田光男, 嶋田恭助, 風間一洋, 佐藤翔
2. 発表標題 arXiv論文に対するTwitterでの言及行動タイプに関する予備調査
3. 学会等名 第34回人工知能学会全国大会
4. 発表年 2020年

1. 発表者名 吉田光男
2. 発表標題 COVID-19流行下でのプレプリントツイートに関する予備調査
3. 学会等名 第11回横幹連合コンファレンス
4. 発表年 2020年

1. 発表者名 小林和央, 風間一洋, 吉田光男, 大向一輝, 佐藤翔, 桂井麻里衣
2. 発表標題 パーストを用いた論文の特性分析
3. 学会等名 第16回Webインテリジェンスとインタラクション研究会
4. 発表年 2020年

1. 発表者名 嶋田恭助, 風間一洋, 吉田光男, 大向一輝, 佐藤翔, 桂井麻里衣
2. 発表標題 Twitter上のarXivプレプリントに関する学術情報流通のキーパーソンの特性分析
3. 学会等名 第16回Webインテリジェンスとインタラクション研究会
4. 発表年 2020年

1. 発表者名 豊島秀典, 吉田光男, 梅村恭司
2. 発表標題 Twitterでの学術情報流通におけるネットワーク作成法とユーザ重要度の関係
3. 学会等名 第16回Webインテリジェンスとインタラクション研究会
4. 発表年 2020年

1. 発表者名 西岡千文, 亀田堯宙, 佐藤翔
2. 発表標題 日本の学術出版物における引用データのオープン化の現状分析
3. 学会等名 情報処理学会 第120回人文科学とコンピュータ研究発表会
4. 発表年 2019年

1. 発表者名 小林和央, 風間一洋, 吉田光男, 大向一輝, 佐藤翔
2. 発表標題 学術情報検索における閲覧論文の文献種別による分析
3. 学会等名 第12回Webとデータベースに関するフォーラム
4. 発表年 2019年

1. 発表者名 嶋田恭助, 風間一洋, 吉田光男, 佐藤翔
2. 発表標題 Twitterで言及されるarXiv論文の特性分析
3. 学会等名 第12回Webとデータベースに関するフォーラム
4. 発表年 2019年

1. 発表者名 小林和央, 風間一洋, 吉田光男, 大向一輝, 佐藤翔
2. 発表標題 学術情報検索における閲覧論文の分析
3. 学会等名 IDRユーザフォーラム2019
4. 発表年 2019年

1. 発表者名 吉田光男, 菊地真人, 梅村恭司
2. 発表標題 科学ニュース記事で取り上げられる学術雑誌に関する予備調査
3. 学会等名 第12回データ工学と情報マネジメントに関するフォーラム
4. 発表年 2020年

1. 発表者名 嶋田恭助, 風間一洋, 吉田光男, 佐藤翔
2. 発表標題 Twitter上のarXivからの学術情報流通に関する分析
3. 学会等名 第12回データ工学と情報マネジメントに関するフォーラム
4. 発表年 2020年

1. 発表者名 小林和央, 風間一洋, 吉田光男, 大向一輝, 佐藤翔, 桂井麻里衣
2. 発表標題 学術情報検索における検索語を用いた論文の特性分析
3. 学会等名 第12回データ工学と情報マネジメントに関するフォーラム
4. 発表年 2020年

1. 発表者名 豊島秀典, 廣中詩織, 吉田光男, 梅村恭司
2. 発表標題 引用関係とTwitter言及による関係をもとにした論文ネットワークの近傍論文の比較
3. 学会等名 第18回Webインテリジェンスとインタラクション研究会
4. 発表年 2022年

1. 発表者名 宇川徹, 風間一洋, 吉田光男, 大向一輝, 佐藤翔, 桂井麻里衣
2. 発表標題 学術情報の閲覧時系列の素性に基づくユーザモデルによるランキング手法
3. 学会等名 WebDB夏のワークショップ2022
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	風間 一洋 (Kazama Kazuhiro) (60647204)	和歌山大学・システム工学部・教授 (14701)	
研究分担者	佐藤 翔 (Sato Sho) (90707168)	同志社大学・免許資格課程センター・准教授 (34310)	
研究分担者	桂井 麻里衣 (Katsurai Marie) (70744952)	同志社大学・理工学部・准教授 (34310)	
研究分担者	大向 一輝 (Ohmukai Ikki) (30413925)	東京大学・大学院人文社会系研究科(文学部)・准教授 (12601)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------