

令和 5 年 6 月 27 日現在

機関番号：32699

研究種目：基盤研究(C)（一般）

研究期間：2017～2022

課題番号：17K00455

研究課題名（和文）共引用ネットワークの拡張と特性分析に基づく文献検索手法の開発

研究課題名（英文）Developing a Document Retrieval Technique by Enlarging and Analyzing Co-citation Networks

研究代表者

江藤 正己（Eto, Masaki）

学習院女子大学・国際文化交流学部・准教授

研究者番号：10584807

交付決定額（研究期間全体）：（直接経費） 3,500,000円

研究成果の概要（和文）：本研究では、共引用ネットワークの拡張と特性分析に基づく検索手法の開発をおこなった。具体的には、共引用ネットワークを拡張することが文献群の自動的な分類において有効であることや段落に含まれる引用の数に着目して解析することの可能性を明らかにした。また、検索実験の環境を整備し、共引用ネットワークの拡張や特性分析に関する基礎的アイデアの効果を本格的な検索実験により検証した。

研究成果の学術的意義や社会的意義

共引用ネットワークに基づく検索には、「キーワード検索では見つけられない適合文献を発見できる」という長所が存在する。本研究では、その長所を「ネットワークの拡張」と「ネットワークの特性分析」の二つの観点から強化した検索手法の開発をおこなった。本研究で開発された検索手法を用いることで、キーワード検索や従来の共引用ネットワークに基づく手法では見つけることのできなかつた適合文献を新たに得ることができるようになった。

研究成果の概要（英文）：This research developed a novel document retrieval technique by enlarging and analyzing co-citation networks. Specifically, this research clarified the performance of document clustering using enlarged co-citation networks, and the possibility of analyzing paragraphs containing many references. In addition, this research prepared experimental environments that enabled large-scale information retrieval experiments, and empirically explored the effectiveness of the basic ideas of enlarging and analyzing co-citation networks.

研究分野：情報検索

キーワード：情報図書館学 情報システム 情報検索 共引用 引用索引 文脈情報 ネットワーク 情報推薦

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

次世代型情報検索システムでは、キーワード検索では見つけられない適合文献をいかにみつけるかが重要な鍵となる。そして学術文献の検索システムでは、「引用関係の利用」が発展の方向性の一つである。特に近年は、大規模な引用データをコンピュータにより高速処理することが可能になってきたため、計量書誌学と情報検索の組み合わせに注目が集まっている。

本研究は、申請者が研究代表者として進めてきた二つの科学研究費補助金などによる研究(「共引用関係を利用した文献検索手法の高度化」と「高度な共引用ネットワークに基づく文献検索手法の開発」)いずれも若手 B。以下、これらを「先行研究」とする)の成果をさらに進展させるものである。先行研究の成果の一つに、共引用ネットワーク拡張のための要素がある(図1)。従来の古典的な検索手法では、文献 A と共引用関係にある文献 B のみが検索結果として出力されていた。先行研究では、多段階化された共引用関係(文献 B とさらに共引用関係にある文献)や、粗い共引用関係(文献 X と出現語が類似した文献 Z が引用している文献)が、より多くの適合文献を得るための要素として有力なことを明らかにしている。

また、先行研究では、引用がおこなわれた文章の文脈の解析を解析し、その結果をもとにネットワークの特性を分析することも試みている。例えば、図1において文献 C をネットワークに組み入れようとする際、文献 Y の本文の文脈を解析し、文献 B と C が「強いつながり(図2の文献 B と C1 のように同一段落で引用されている)」の場合は組み込み、「弱いつながり」の場合は除外することで、ネットワーク中の不適合文献の数を抑えられることを明らかにしている。

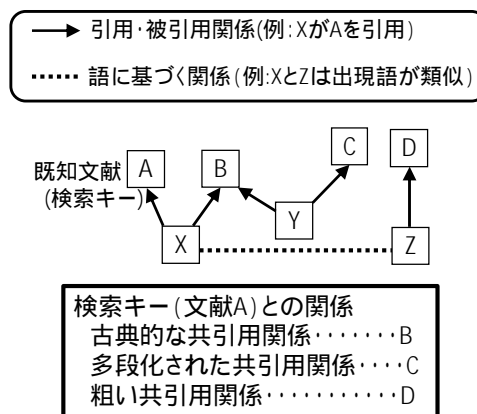


図1 ネットワーク拡張の要素

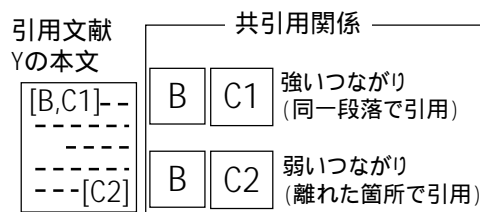


図2 引用文脈に基づく共引用関係の強弱

2. 研究の目的

本研究の目的は、共引用ネットワークを利用した検索において、その長所「キーワード検索では見つけられない適合文献を発見できる」を強化した新たな検索手法を開発することである。具体的には、共引用ネットワークを拡張すること、及びその特性を分析することにより、検索手法の性能向上を図る。

3. 研究の方法

まず、共引用ネットワークの拡張に関して、「粗い共引用関係」に基づく拡張方法が、クラスタリングにおいても機能することを確認した。次に、共引用ネットワークの特性分析に関して、「多数の引用がおこなわれた段落における共引用関係」の強度について、調査した。さらに、検索実験をおこなう環境を整備し、共引用ネットワークの拡張や特性に関する基礎的アイデアの効果を本格的な検索実験により検証した。

4. 研究成果

(1) 共引用ネットワークの拡張

共引用ネットワークの拡張に関して、「粗い共引用関係」に基づく拡張方法が、クラスタリングにおいても機能することを確認した。「粗い共引用関係」とは図1で示すような、起点となる既知文献(文献 A)と「その起点文献を引用している文献に含まれる語を手がかりにして語に基づく検索をおこない、そこで得られた文献が引用している文献」(文献 D)との関係である。「粗い共引用関係」は、関連性の強度は古典的な共引用関係(文献 A と文献 B の関係)よりも弱いものの、従来の共引用関係では得ることの出来ない関連性をとらえることができる。先行研究では、「粗い共引用関係」をネットワークに取り込むことが文献検索において有効であることを明らかにしている。本研究では、この粗い共引用関係に基づくネットワークの拡張の有効性をさらに調査するために、文献群のクラスタリングにおける効果を検証した。

具体的には、「従来の共引用関係に基づいて算出した類似度のみによってクラスタリングをおこなう手法」と「従来の共引用関係に基づいて算出した類似度、粗い共引用関係に基づいて算出した類似度を組み合わせてクラスタリングをおこなう手法」を比較することによりおこな

た。比較実験は、同一の文献群を対象に適用し、その性能を比較することによりおこなった。その結果、従来の共引用関係と粗い共引用関係を組み合わせた手法の方がクラスタリングの性能が高いことが示された。このことから、粗い共引用関係に基づく共引用ネットワーク拡張の有効性を、文献検索の応用的な範疇であるクラスタリングにおいても確認することができた。

(2) 共引用ネットワークの特性分析

共引用ネットワークの特性分析に関して「多数の引用がおこなわれた段落における共引用関係」の強度について、調査をおこなった。先行研究では、同一の段落で引用された論文同士の間に見られる共引用関係（図2における文献BとC1）は、すべて等しく強いつながりがあるものとして、ネットワークの強度の計算に用いられてきた。

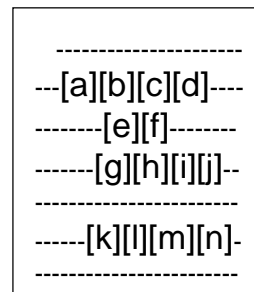


図3 一段落に多数の引用が存在する例

しかしながら、図3で示すような、一つの段落内で多くの論文に対して引用がなされるような場合においては、そこで生じる共引用関係のつながりが弱い可能性があることが先行研究において指摘されていた。本研究では、この問題をさらに検証するため、二つの調査をおこなった。一つは「多数の引用がおこなわれた段落における共引用関係」の数がどれくらい存在するかの調査であり、もう一つが「多数の引用がおこなわれている段落における共引用関係の強度」と「少数の引用がおこなわれている段落における共引用関係の強度」の比較である。具体的には、CITRECのデータセットに含まれる共引用関係の情報を対象とし、段落に含まれる引用数別に、共引用関係にある論文同士の類似度を計算し、引用数と共引用関係の強度の関わりを分析した。

調査の結果、「多数の引用がおこなわれた段落における共引用関係」の数は少なくないことが示された。また、「多数の多数の引用がおこなわれている段落における共引用関係」の強度は、「少数の引用がおこなわれている段落における共引用関係」の強度よりも弱い傾向がみられることが明らかになった。これらのことから、段落に含まれている引用の数を考慮し、それらを共引用ネットワークにおけるリンク強度の計算に反映させることが有効である可能性が示唆された。

(3) 文献検索における共引用ネットワークの拡張や特性分析に関する基礎的なアイデアの検証 実験環境の構築

先行研究は、文献検索における共引用ネットワークの拡張や特性分析に関する基礎的なアイデアを初期的な検索実験においてしか検証していなかった。より本格的な検索実験を可能とするために、検索実験の環境を整備することをおこなった。

まず、従来よりも現実の環境に近いデータセットを作成した。従来は、検索実験をおこなう際に、簡易的な学術論文データセット（検索された論文が適合しているか否かを判定する時に、論文の内容を示す索引用の語に基づいて機械的に計算された類似度を利用する）のみを用いていた。これを本格的なデータセット（人間が適合しているか否かを判定するような要素を取り入れた、より現実の状況に近いもの）も用いることができるよう実験環境を整備した。

次に、開発検索システムの実装にグラフデータベースを導入した。これにより、グラフやネットワークに関する様々なアルゴリズムを効率的に利用できるようになり、効果的な検索手法開発をおこなえるようになった。

共引用ネットワークの拡張や特性分析に関するアイデアの検証実験

共引用関係の多段階化（図1における文献D）と図2で示した引用文脈に基づいて共引用関係の強弱（図2）を組み合わせた検索手法を、先行研究の実験よりも、広く深い観点から検証した。この検証においては、先行研究でも用いられてきた簡易的な学術文献データセットに加え、新たに構築した本格的なデータセットも用いた。また、基礎的なアイデアで構築された共引用ネットワークに適したランク付けアルゴリズムを多角的に検討するため、実験に用いるアルゴリズム種類数を1から5に増やした。二つのデータセットと五つのアルゴリズムを用いた検索実験により、共引用ネットワークの拡張や特性分析に関するアイデアの効果について考察をおこなった。

また、粗い共引用関係に基づく検索手法をキーワード検索と比較する検証をおこなった。先行研究においては、粗い共引用関係に基づく検索手法を評価する際、従来の共引用関係や多段階化された共引用関係に基づく検索手法との比較をおこない、より多くの適合文献を見つけられていることを確認する作業を主におこなっていた。しかしながらこの評価方法は、粗い共引用関係により得られた新たな適合文献とキーワード検索で得られた適合文献に重なりがあるか否かに

については分析できておらず、本研究の目的である「キーワード検索では見つけられない適合文献を発見できる」の観点において検索性能を十分に評価できていない問題があった。そこで、粗い共引用関係に基づく検索で見つけられる適合文献とキーワード検索で見つけられる適合文献とを検索実験により比較した。その結果、粗い共引用関係に基づく検索手法がキーワード検索では見つけることのできない適合文献を多数発見できていることが明らかとなった。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 1件）

1. 著者名 Masaki Eto	4. 巻 56
2. 論文標題 Extended co-citation search: Graph-based document retrieval on a co-citation network containing citation context information	5. 発行年 2019年
3. 雑誌名 Information Processing and Management	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.ipm.2019.05.007	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 -

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 2件）

1. 発表者名 Masaki Eto
2. 発表標題 Strength of Co-citation Linkages Observed in Paragraphs with Many References
3. 学会等名 METRICS 2018: Workshop on Informetric and Scientometric Research (SIG/MET) (国際学会)
4. 発表年 2018年

1. 発表者名 Masaki Eto
2. 発表標題 Can Rough Co-citation Improve the Performance of Co-citation Clustering?
3. 学会等名 ACM/IEEE Joint Conference on Digital Libraries in 2018 (JCDL 2018) (国際学会)
4. 発表年 2018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
---------------------------	-----------------------	----

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
---------	---------