

平成30年6月25日現在

機関番号：32699

研究種目：若手研究(B)

研究期間：2014～2017

課題番号：26730163

研究課題名(和文) 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

研究課題名(英文) Developing a Document Retrieval Technique Using Improved 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) 2,600,000円

研究成果の概要(和文)：本研究では、共引用ネットワークを語のつながりがある文献(出現語が類似した文献)を介して高度化させる、新たな検索手法を開発した。具体的には、共引用ネットワークを適切に拡張するための手がかりとして、「周辺文献」や「粗い共引用関係」を考案した。さらに、これらの手がかりに基づいて拡張させた共引用ネットワークによる検索手法を、従来の共引用関係に基づくネットワークによる検索手法と比較する実験をおこない、拡張をおこなった手法の検索性能の方が高いことを確認した。

研究成果の概要(英文)：This research developed a novel document retrieval technique based on co-citation networks that were improved by word-based linkages. Specifically, this research proposed two ideas: 'satellite documents' and 'rough co-citation' as keys to effectively enlarge co-citation networks. Results of the information retrieval experiments showed that the proposed retrieval methods, which used co-citation networks enlarged by these ideas, outperformed a baseline retrieval method using original 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

1. 研究開始当初の背景

次世代型の文献検索システムでは、キーワード検索は当然のものとして、「それを補完する検索機能をいかに充実させるか」が重要な要素である。そして、学術文献検索の検索システムでは、「キーワード検索では見つけられない適合文献を発見できる」という長所を持つ、引用関係を利用した検索をさらに充実させることが有力な発展の方向性の一つといえる。

本研究は、研究代表者がこれまでにこなってきた科学研究費補助金による「共引用関係を利用した文書検索手法の高度化」(若手研究B:代表者:江藤正己、以下「先行研究」とする)の研究成果を発展させるものである。共引用関係とは、一つの文献が二つの文献を同時に引用していた場合の、二つの被引用文献間の関係のことを指す。この関係を用いた検索では、検索キーとなる既知文献を入力し、「既知文献と共引用関係にある文献群」が「類似文献」として出力される。たとえば、図1では、文献Bが文献Aと文献Cを引用しているため、文献Aをシステムに入力した場合、文献Cが出力される。先行研究は、このような共引用関係に基づく検索を発展させるものであった。

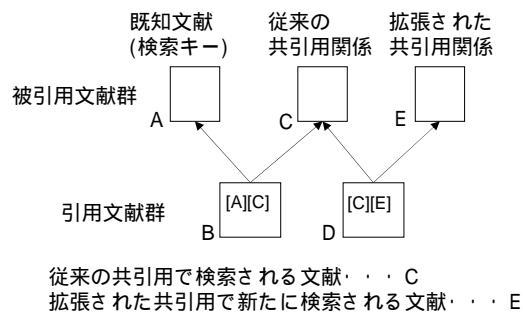


図1 共引用関係の多段階化

先行研究の主な成果として、次の二つのことが明らかになっている。一つ目は、図1の文献Eのように共引用関係の概念を多段階に押し広げることで、従来よりも多くの適合文献が得られ再現率が向上することである。二つ目は、共引用関係を拡張する際に、その引用の文脈を機械処理で解析して利用することが精度の面で有用であることである。

2. 研究の目的

本研究の目的は、引用関係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。より具体的には、多段階化された共引用関係をネットワークモデルでとらえ、そのネットワークを適切に拡張させることによって高度化し、検索手法の性能向上を図る。

3. 研究の方法

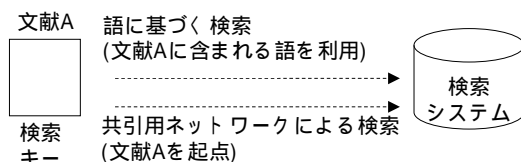
本研究では、共引用ネットワークを語のつながりがある文献(出現語が類似した文献)を介して拡張させることにより、その検索性能の向上を試みた。

まず語と共引用ネットワークの組み合わせの有効性を確認したうえで、拡張のための手がかりとして、「周辺文献」や「粗い共引用関係」を考案した。さらに、考案した手がかりに基づいて拡張をおこなったネットワークによる検索手法と、従来の共引用関係に基づくネットワークなどによる検索手法とを比較分析し、考案した方法の有効性を確認した。比較は、テストコレクションを用いた検索実験によって実証的におこなった。また、検索実験におけるネットワーク中の文献のランク付けには、グラフ理論に基づくアルゴリズムなどを用いた。

4. 研究成果

(1) 語と共引用の組み合わせの有効性

共引用ネットワークと語の組み合わせの有効性を確認した。図2は、文献Aに含まれる語をキーとした検索結果のランキングと文献Aを起点とした共引用ネットワークによる検索結果のランキングを示している。図2で示すような同じ文献を検索キーとした場合の二つの検索結果のランキングでは、その順位を組み合わせる新たなランキングを作り、組み合わせ前のランキングと比較することで、組み合わせの効果を確認することができる。



順位	語に基づく検索	順位	共引用ネットワークによる検索
1	文献ID#111	1	文献ID#127
2	文献ID#121	2	文献ID#111
3	文献ID#321	3	文献ID#425

図2 検索結果ランキングの組み合わせ

そこで、「語に基づく検索と共引用ネットワークによる検索を組み合わせる手法」を「語に基づく検索と単純な共引用関係による検索を組み合わせる手法」や「(組み合わせをおこなわない)語だけに基いた検索手法」と比較する検索実験をおこなった。その結果、「語に基づく検索と共引用ネットワークによる検索を組み合わせる手法」が、他の二つの手法よりも検索性能が優れている傾向がみられた。このことから、共引用ネットワークの拡張において、語との組み合わせをおこなうことが有効であることが明らかとなった。

(2) 共引用関係の周辺文献を利用した検索
共引用ネットワークを適切に拡張する手がかりとして、共引用関係にある文献の「周辺文献」を考案した。検索キーを文献Aとした場合の共引用ネットワークと周辺文献の例を図3で示す。図3において、文献Cに含まれる語をキーとした全文検索で得られた文献群が、文献Cの周辺文献となる。周辺文献は、さらに、ネットワークに含まれていない新規の文献(I, J, K)と既にネットワークに存在していた既存文献(E, F, G)に分けられる。周辺文献を共引用ネットワークに取り込んで拡張をおこなう際、新規文献については、新たなノードの作成やエッジの追加に、既存文献は、エッジの追加や重みの再計算に利用される。

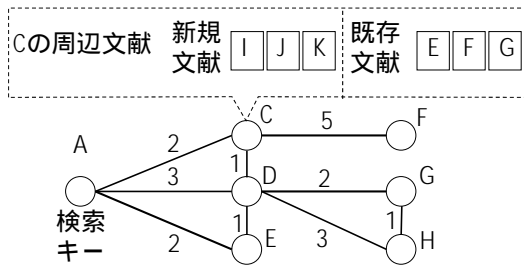


図3 共引用ネットワークと周辺文献

この周辺文献を手がかりとしたネットワークの拡張方法を評価するために、「従来の共引用関係に基づくネットワークによる検索手法」と「周辺文献に基づいて拡張した共引用ネットワークによる検索手法」とを検索実験により比較分析した。その結果、後者の検索性能の方が優れている傾向がみられ、共引用ネットワークを拡張させる手がかりとして「周辺文献」が有効であることが明らかとなった。

(3) 粗い共引用関係を利用した検索
粗い共引用

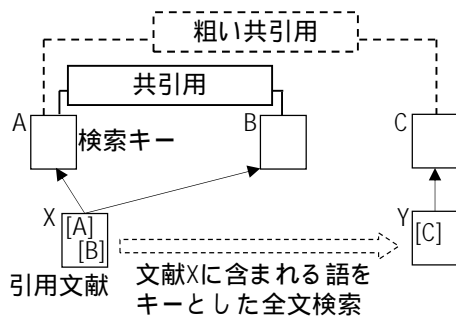


図4 粗い共引用関係

共引用ネットワークを適切に拡張させる別の手がかりとして「粗い共引用関係」を考案した。図4において、粗い共引用関係の例を示す。この図では、文献AとBは文献Xに引用されており、両者の間に共引用関係のみ

ることができる。また、文献Yは文献Xに含まれる語をキーとした全文検索で得られた文献であり、文献Xと同じような出現語をもつ文献である。

文献Aと文献Bの間にある共引用関係が、両者の何らかの関連性を示すものであると考えられる理由は、両者が一つの文献という文脈において、引用されているためである。このことをふまえた場合、同じような出現語をもち文脈が似ている文献Xと文献Yのそれぞれから引用されている文献Aと文献Cの間にも、なんらかの関連があると考えられることができる。本研究では、このような文献Aと文献Cの関係性を、「粗い共引用」と定義した。粗い共引用関係は、ネットワークに取り込まれる際に、ノードの追加、およびエッジの追加や重みの再計算に利用される。

なお、この粗い共引用関係は、従来の共引用関係よりも、関連性の度合いが弱いことが想定される(図4の場合、共引用関係ABと粗い共引用関係ACでは、ACの方がABよりも関連の度合いが弱い)。そのため、粗い共引用関係をネットワークに組み込む際には、減衰係数を設定し(例えば、粗い共引用関係の関連度の強さを従来の共引用関係の強さの半分と想定する場合、係数を0.5とするなど)、粗い共引用関係の回数に減衰係数を乗じてエッジの重みを計算する。

この粗い共引用関係を手がかりとしたネットワークの拡張方法を評価するために「従来の共引用関係に基づくネットワークによる検索手法」と「粗い共引用関係に基づいて拡張した共引用ネットワークによる検索手法」とを検索実験により比較分析した。その結果、後者の検索性能の方が優れている傾向がみられ、共引用ネットワークを拡張させる手がかりとして、「粗い共引用関係」が有効であることが明らかとなった。

粗い共引用関係の応用的な利用

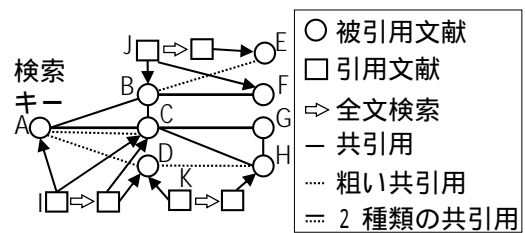


図5 応用的な粗い共引用ネットワーク

前述の手法は、「検索キーとなる文献と粗い共引用関係にある文献群」を共引用ネットワークに取り込むのみであった。たとえば図5の場合、この手法は、文献Aを引用している文献Iに含まれる語を利用して、粗い共引用関係を求め、それらをネットワークに取り込むのみである。

そこで、共引用ネットワークをさらに拡張するための、粗い共引用関係の応用的な利用

方法を考案した。具体的には、「検索キーと共引用関係にある文献」を引用している文献（図5文献J）や「検索キーと粗い共引用関係にある文献」を引用している文献（図5文献K）に含まれる語を利用して、さらに粗い共引用関係を求め、ネットワークを拡張する方法である。

「 の手法」と「応用的な粗い共引用関係に基づいて拡張したネットワークによる検索手法」とを検索実験により比較分析した結果、後者の検索性能の方が優れている傾向がみられた。このことから、共引用ネットワークを拡張する際の粗い共引用関係の応用的な利用方法として、考案した方法が有効であることが明らかとなった。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔学会発表〕(計 4件)

Masaki Eto, Increasing Source Documents of Rough Co-citation to Expand Co-citation Networks for Scientific Paper Searches, METRICS 2017: Workshop on Informetric and Scientometric Research (SIG/MET), 2017年

Masaki Eto, Rough Co-citation as a Measure of Relationship to Expand Co-citation Networks for Scientific Paper Searches, Proceedings of the 79th ASIS&T Annual Meeting, 2016年,
DOI:10.1002/pr2.2016.14505301131

Masaki Eto, Incorporating Satellite Documents into Co-citation Networks for Scientific Paper Searches, Proceedings of the Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL) co-located with the Joint Conference on Digital Libraries 2016, 2016年,
<http://ceur-ws.org/Vol-1610/paper4.pdf>

Masaki Eto, Combination Effects of Word-based and Extended Co-citation Search Algorithms. Proceedings of the 15th ACM/IEEE-CE Joint Conference on Digital Libraries, 2015年
DOI:10.1145/2756406.2756957

〔図書〕(計 1件)

江藤正己, 勁草書房, 利用に基づく情報の組織化(『図書館は市民と本・情報をむすぶ』), 372 (137-146), 2015年

6. 研究組織

(1)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