

令和元年6月6日現在

機関番号：32665

研究種目：基盤研究(C)（一般）

研究期間：2016～2018

課題番号：16K00445

研究課題名（和文）クラウドソーシングと集合知を活用した文献探索手法の研究開発

研究課題名（英文）literature survey using crowdsourcing and collective knowledge

研究代表者

韓 東力（HAN, Dongli）

日本大学・文理学部・教授

研究者番号：10365033

交付決定額（研究期間全体）：（直接経費） 2,800,000円

研究成果の概要（和文）：これまでの文献探索は、論文中のキーワードを利用したものが中心となっている。この方法は検索スキルを必要とする上、検索結果に関連性の低いものが多く含まれてしまうという欠点がある。CiNiiやGoogle Scholar、CiteSeerなどのように文献間の参照関係を辿ることが可能な文献検索サービスもあるが、論文間の参照・被参照の関係だけを頼りにしているため、効率的な探索ができかねる。本研究では、論文間参照ネットワークを利用した文献探索に焦点を当て、論文間参照理由を参照ネットワークに付与することにより、起点論文の関連文献を効率よく探索する手法を開発し、その効果を評価実験により検証する。

研究成果の学術的意義や社会的意義

参照ネットワークを辿ることにより関連文献探索を試みる既存研究が複数存在するが、いずれも参照・被参照以外の論文間関係を考慮しておらず、参照ネットワークが広ければ広いほど、文献選定の手間が増えていくため効率的な探索ができない。実験を通して本研究手法の有効性が確認できれば、たとえば「起点論文で用いられている基礎理論がより詳細に紹介されている文献」のようなきめ細かい探索ニーズにも効率的に対応できるようになり、研究者の研究活動に大きな手助けとなることが期待できる。

研究成果の概要（英文）：Literature survey is the first step of scientific research. However, this process could be quite time-consuming. Previous works aiming to automatically address this issue employ textual similarity or reference-relation between papers, while neither of which is flexible enough for context-specific demands. In this paper, we introduce the idea of using citation-reasons to narrow down the search range for relevant papers. We first propose a method to predict citation-reasons between scientific papers with machine-learning techniques. However, as the machine-learning method seems not accurate enough according to a subject experiment, we then propose another strategy to annotate citation-reasons between papers in a crowdsourcing manner. The experimental results have shown the effectiveness of our proposal.

研究分野：知能情報学

キーワード：文献探索 機械学習 クラウドソーシング 集合知

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

これまでの文献探索は、起点論文中のキーワードを利用したものが中心となっている。この方法は検索スキルを必要とする上、検索結果に関連性の低いものが多く含まれてしまうという欠点がある。CiNiiやGoogle Scholar、CiteSeerなどのような文献間の参照・被参照関係を辿ることが可能な文献検索サービスの場合、最初から探索の範囲を絞り込むことができるが、論文間の参照・被参照関係以外の手がかりがなく、参照ネットワークが広ければ広いほど、文献選定の手間が増えていくため効率的な探索ができない(図1参照)。

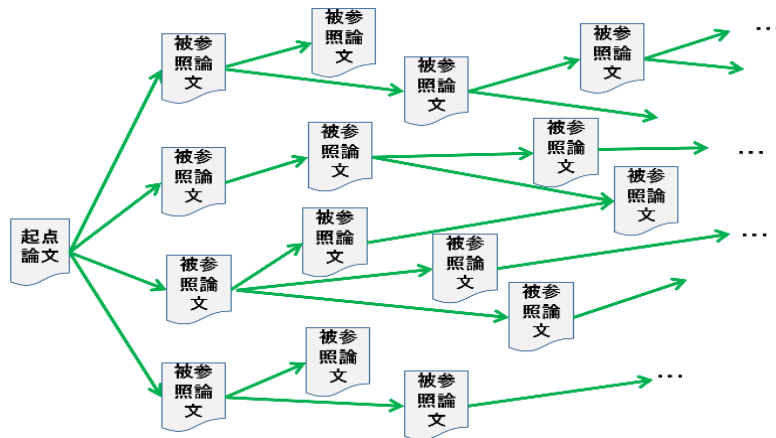


図1 参照ネットワークのイメージ

2. 研究の目的

本研究では、このような参照ネットワークを利用した文献探索に焦点を当て、文献探索者の探索目的に応じた、全体の参照ネットワークから切り取った部分ネットワークを辿ることにより、起点論文の関連文献を効率よく探索する手法を開発する(図2参照)。

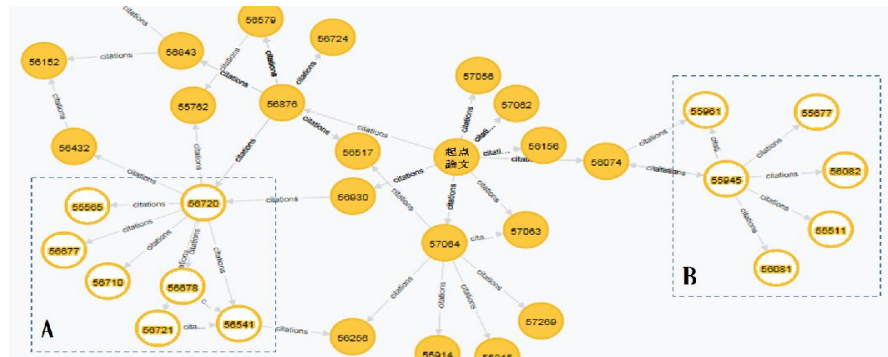


図2 論文探索効率化のイメージ

3. 研究の方法

論文間の参照理由とは、起点論文がどのような理由で他の論文(以下「被参照論文」と呼ぶ)を参照しているのかを具体化したもので、「背景紹介」や「データ利用」など10種類前後が定義されている(図3参照)。これらの参照理由が組み合わせることで、さらに複合化してさまざまな参照関係を解釈できる。たとえば、起点論文で述べられた研究手法の正当性の裏付けとして被参照論文を参照している場合や、起点論文で用いられている基礎理論が詳細に記述されている被参照論文を参照している場合など、様々な参照理由が考えられる。そこで、文献探索者の探索目的を論文間参照理由に置き換えられれば、参照理由があらかじめ付与されている参照ネットワークを探索目的で検索し、マッチングした部分ネットワークのみを辿ることにより、起点論文の関連文献を効率よく探索できるのではないかと考えた。この考え方では論文間参照理由の付与精度がカギとなるため、機械学習による自動付与の結果に加え、より正確性が高いと言われているクラウドソーシングと集合知の考え方も導入した。

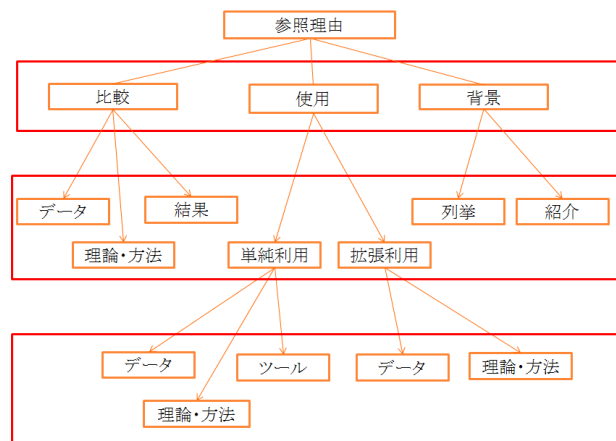


図3 参照理由分類階層

4. 研究成果

本研究では、論文間参照ネットワークを利用した文献探索に焦点を当て、論文間参照理由を参照ネットワークに付与することにより、起点論文の関連文献を効率よく探索する手法を開発し、

その効果を実験により検証することを目的としている。平成 28 年度は機械学習による論文間参照理由の自動判定手法を開発し、平成 29 年度はその判定精度の向上を目指し、クラウドソーシングを導入した。そして最終年度の平成 30 年度は、平成 29 年度に続き論文間参照ネットワークの構築と文献探索者の探索目的を参照理由に置き換える仕組みを設計した上で、論文間参照理由にもとづく文献探索システムを構築した(図 4 参照)。また、探索範囲内において起点論文の関連文献を探索者自身の判断により行う必要があるが、探索者の負担を軽減するために作業を円滑に行ってもらえるように、パスの強調表示やノード(論文)の情報表示など、一連の補助機能も併せて開発した。システムでは、関連性の低い文献や探索目的にそぐわない文献の一部が切り捨てられるため、文献探索の効率向上が確認された。最終年度では本研究の成果を内外の学会で発表しながら、社会に向けての情報発信を積極的に行った。

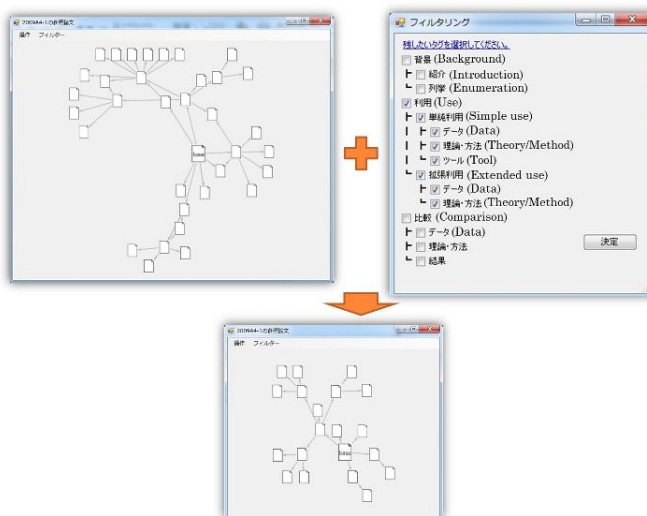


図 4 システムインタフェース

5 . 主な発表論文等

[雑誌論文](計 1 件)

Yingying Lao, Ryosuke Kasai, Tong Wu, Dongli Han, Corpus-generation Support System Based on News Websites with URL Extraction and News Selection Techniques, International Journal of Digital Content Technology and its Applications. Vol.13. 2019. pp.144-153. (Reviewed)

[学会発表](計 9 件)

Dongli Han, Literature Survey Using Crowdsourcing, The 2019 International Conference on Knowledge Engineering and Applications. 2019. (招待講演)

Yilun Wei, Yingying Lao, Yudai Sato, Dongli Han, Product-review Classification Combining Multiple Clustering Algorithms. The 2019 International Conference on Knowledge Engineering and Applications. 2019.

Yingying Lao, Takumi Hosokawa, Tong Wu, Dongli Han, Elucidation of the influence of the structure of lyrics on the ease of understanding lyrics. The 4th Asia Pacific Corpus Linguistics Conference. 2018.

Yingying Lao, Tomoki Aoyama, Dongli Han, Exploitative-enterprise Diagnosis Based on Users' Values. 第 17 回情報科学技術フォーラム. 2018.

Dongli han, Ayato Inoue, Efficiency Improvement of Literature Survey Based on Citation-Reason Visualization. 2017 International Conference on Knowledge Engineering and Applications. 2017.

Dongli Han, Ayato Inoue, Yousuke Yajima, Zhengliang Sun, Rina Funaki, Citation-reason Annotation Using Crowdsourcing. 2017 International Conference on Artificial Intelligence: Techniques and Applications. 2017.

Dongli Han, Ryo Fukuoka, Genki Wakabayashi, Taro Shimizu, Shinnosuke Masuda, Comparison of Multiple Recommendation Methods of Similar Onomatopoeia. The 12th International Conference on Computer Science and Education. 2017.

Dongli Han, Takehiro Ishibashi, Kyosuke Takanami, Ryo Shinozaki, A Method to Extract Size-feelings from Customer-reviews. The 8th International Conference on Information. 2017.

Dongli Han, Takahiro Ohno, A Case Study on Experimental-data validation for Natural Language Processing, The 11th International Conference on Computer Science and Education. 2016.

〔図書〕(計 件)

〔産業財産権〕

出願状況(計 件)

名称：
発明者：
権利者：
種類：
番号：
出願年：
国内外の別：

取得状況(計 件)

名称：
発明者：
権利者：
種類：
番号：
取得年：
国内外の別：

〔その他〕

ホームページ等

6. 研究組織

(1) 研究分担者

研究分担者氏名：

ローマ字氏名：

所属研究機関名：

部局名：

職名：

研究者番号(8桁)：

(2) 研究協力者

研究協力者氏名：

ローマ字氏名：

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、研究者個人に帰属されます。