

学術論文における関連研究の執筆支援のための タスク設計およびデータ構築

成松宏美¹ 小山康平² 堂坂浩二³ 田盛大悟⁴
東中竜一郎¹ 南泰浩² 平博順⁴

¹NTT コミュニケーション科学基礎研究所 ²電気通信大学
³秋田県立大学 ⁴大阪工業大学

{hiromi.narimatsu.eg, ryuichiro.higashinaka.tp}@hco.ntt.co.jp
k1710245@edu.cc.uec.ac.jp, dohsaka@akita-pu.ac.jp, e1b17043@st.oit.ac.jp
minami.yasuhiro@is.uec.ac.jp, hirotooshi.taira@oit.ac.jp

1 はじめに

学術論文を執筆する際、関連する文献の調査および適切な引用は重要である。新しいテーマの研究を始めるときや研究課題に対する新しいアプローチについて考えるとき、過去の研究に立脚し、また、論文を執筆する際には、記述内容の根拠について、参考文献を正しく示すことが求められる。これらは重要な研究活動の一環ではあるものの、分野の広がりや対象とすべき論文の増加から、その調査および執筆を支援することを目的とした研究が行われてきた[1, 2]。従来は研究者による文献調査の量を減らすという観点で、関連研究の要約を目的とした研究が多かったが[1, 3]、近年では大規模な文献データが手に入るようになってきたことから、学術論文の執筆を直接的に支援する引用文献推薦や関連研究の章に書かれる引用文生成の手法が研究されてきた。たとえば、生命科学や生物医学分野の論文の検索エンジンである PubMed¹⁾から取得した文献を主に用い、キーワードだけでなく内容も踏まえて引用すべき文献を推定する手法を提案するものや[4]、関連研究を自動生成する手法[5]、執筆・査読の効率化を目的として、対象となる文に引用が必要かを判定する手法[6]が提案されている。

これらの研究は、いずれも論文執筆の効率化という同じ目標を目指しているものの、それぞれ独立して研究が行われてきた。近年では、論文データベースのAPIの公開によって、論文が容易に手に入るよ

うになり、こうした研究論文を扱うタスクが共有タスクになりつつある[7, 8, 9]が、いずれも異なるタスクを目的としており、統合的な性能評価ができる状態にはない。一方で、論文執筆支援の実用化という観点から考えると、引用すべき文献の推薦、引用すべき文の判定、関連研究の自動生成は連続した研究活動であり、同じデータを対象とした統合的な性能評価ができることが望ましいと考えられる。

そこで、我々は、関連研究の統合的な執筆支援を目的として、必要なタスクおよび同じデータを元にタスクデータを構築する手法を提案する。統合的な執筆支援を目的とすることで、これまでに取り組みがなかったタスクに気づくことができる可能性がある。また、一つのデータを元にタスクデータの構築ができれば、タスク間での同一モデルの適用が検討でき、各タスクの精度の直接比較が可能となるため、タスクの違いや難しさも評価可能になると考えられる。本稿では、関連研究に着目した論文執筆に関わる目標とそれに必要なタスクを定義し、タスクデータの構築方法について述べる。

2 関連研究

論文において関連研究の章は、自らの研究の位置付けを明確化するために重要である。しかし、文献の検索ならびに精読に多くの時間を割かなければならないことから、その効率化が期待されており、多くの研究がなされてきた。

論文執筆支援を目的とした研究の初期において、論文全体を読まなくても論文の要旨および主張を把握できるよう、目的、課題、手法の観点での要約[1]

1) <https://pubmed.ncbi.nlm.nih.gov/>

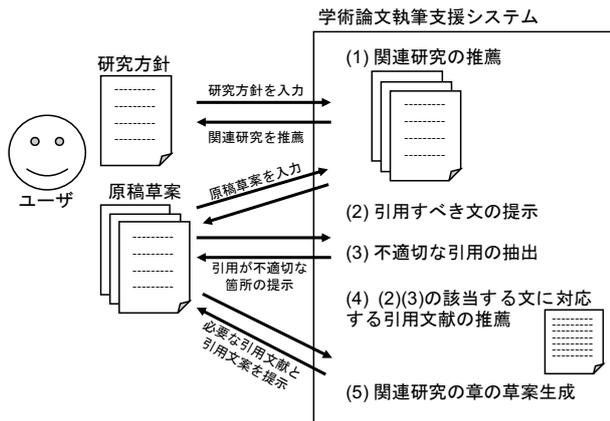


図1 関連研究に着目した統合的な論文執筆支援の流れ

や、該当論文に関する引用文を用いた要約手法が提案されてきた [3]. 学術論文の要約は、一般文書の要約と異なる難しさが指摘されており、今も継続して研究されている [10]. しかし、論文執筆支援という観点では、研究者は別途読むべき論文を探す必要があり、支援としては十分でない。

近年では論文データの取得が容易になってきたことから、引用すべき論文の推薦を目的とした研究も行われている。多くの既存研究は、他でも引用されている権威ある文献を推薦、キーワードの近い文献を推薦する方法が主であったが、近年では他では引用されていない重要な文献や異なる技術用語を使用しているが類似している研究も見つけられるよう、大規模なデータを構築して引用関係の意味ネットワークを考慮した推薦手法も提案されている [11, 4]. これらは関連研究を推薦するという観点において有益であるものの、さらに細かい、引用文に対する引用文献の推薦などは対象とされおらず、本タスクを行うには別途タスクデータの構築が必要となる。

使用データや方法、問題点などのカテゴリに分類して文献を推薦する手法も提案されているが [12], 引用文として記述したものの根拠として、引用すべき文献を提示するタスクとはなっていない。

執筆・査読の効率化を目的とした研究として、引用が必要な文かどうかを判定するタスクもある [6]. この引用要否判定タスクは、執筆において引用漏れを防止する、論文チェックの観点でも重要なタスクではあるが、引用要否判定に続いて、引用すべき文献を推薦するタスクまでつなげて手法の開発や評価ができることが望ましい。従来研究 [6] では、科学論文の大規模データを使用し公開しているものの、

タスク	定義
(1)-1	アブストラクト(目的やアプローチ)を入力としたときに、読むべき文献のリストを提示してくれる(関連研究検索タスク)
(1)-2	ある程度書き終えた論文を入力としたときに、読むべき文献のリストを提示してくれる(被引用文献推薦タスク)
(2)	関連研究の章の各文に対して、引用が必要な文かどうかを当てる(引用要否判定タスク)
(3)-1	引用文献のプールからどの文に対する引用文献かを当てる(引用文献割り当てタスク)
(3)-2	引用文と被引用文献との関係から、引用文献が正しいかどうかを当てる(引用文・被引用文献ペア適正性判定タスク)
(4)	引用文に対応する文献を推薦(引用文献検索および割り当てタスク)
(5)-1	引用文献を関連研究のカテゴリに分類する(引用文献分類タスク)
(5)-2	引用すべき文献の推薦とその引用文の自動生成(被引用文献推薦および引用文生成タスク)
(5)-3	関連研究の章の自動生成(関連研究生成タスク)

表1 タスクの定義

文献の推薦タスクに使用できるデータにはなっておらず、続くフェーズで期待される文献推薦を行うには別途データ収集および構築が必要となる。

さらに、関連研究の章の一部である引用文を被引用文献から自動生成する手法 [5] も提案されているが、データ数が少なく、同一著者の他の文献や、代表的な引用文の書き方などルールベースの手法にとどまっている。

そこで、我々は、関連研究に関わる論文執筆の支援を目的として、連続的かつ必要なタスクを設計し、それらのタスクを統合的に評価できる環境を提供としてタスク用データの構築を行う。

3 タスクの定義

我々の目指す統合的な支援の流れを図1に示す。研究を始める段階では、“自分なりに調べた従来研究を元に課題とアプローチを検討したが他に類似する研究があるかもしれない”と思い、関連研究を探し求める。こうした状況においては、課題とアプローチが記述された研究方針を入力として、関連する文献が推薦されることが望まれる [13](図の(1)). また、原稿執筆の段階では、一通りの草案を書いたが、引用すべき箇所引用がなされなかったり、適切な引用がなされていないことも起こりうる。したがって、引用が漏れている文の提示 [6] や(図の(2)), 異なる文献が引用されている箇所特定(図の(3)), さらに適切な引用文献の提示(図の(4))が期待される。さらには、“一通り研究を終えたものの引用すべき何かが漏れている可能性がある”という状況においては、引用すべき文献と共に、該当する研究に対してどんな観点での引用が望ましいか

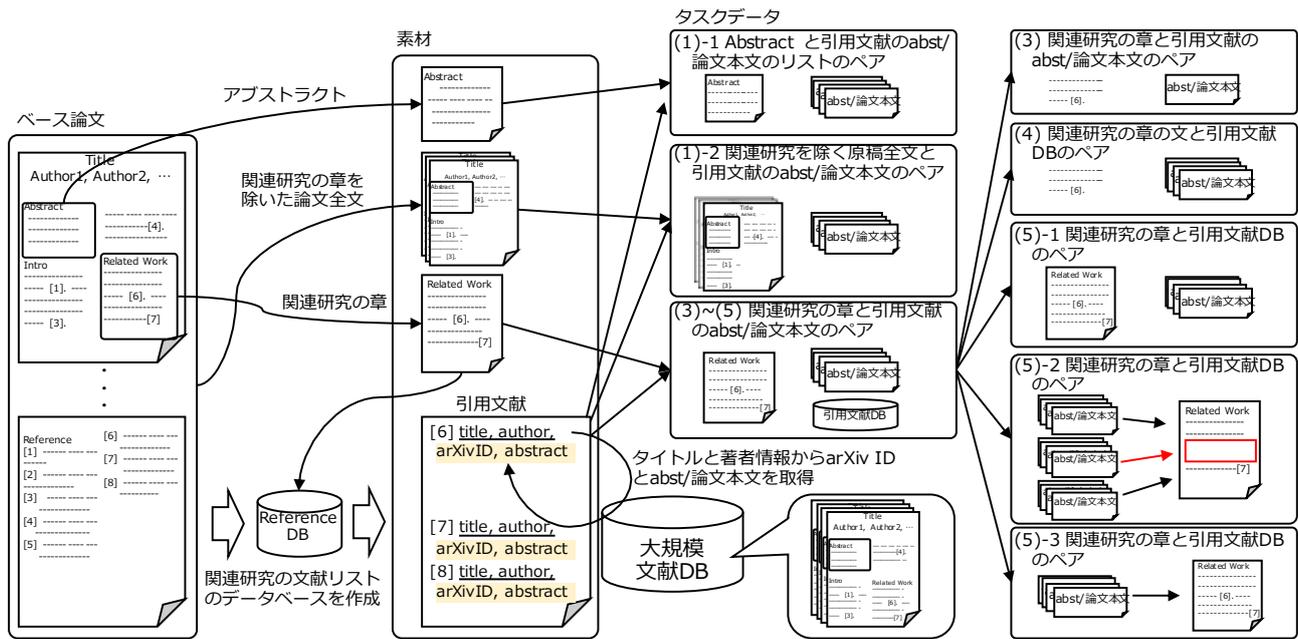


図 2 タスクデータの生成

の位置付けを整理した文章が提示される [5] ことが期待される (図の (5)). そこで、我々の目標に必要なタスクを表 1 のように定義した。以降、入力として用いる研究方針や原稿草案のことを、ベース論文と呼ぶこととする。

4 データ構築

データの構築方法について、図 2 を用いて説明する。論文がたくさん収録されているデータベースを大規模文献 DB と呼ぶ。

はじめに、ベース論文と大規模文献 DB から、タスクデータに用いる素材、すなわちアブストラクト、関連研究の章を除いた論文全体、関連研究の章を抽出する。また引用文献については、タイトルと著者名を用いて大規模文献 DB と照合し、タイトル、著者、arXiv の ID およびアブスト/論文本文をペアにしておく。

最後に素材を元に、タスクデータを生成する。生成方法の全体像を図に示し、データの詳細を以下に述べる。

(1) 読むべき文献の推薦

ベース論文のアブストラクトと引用文献のアブストラクト/論文本文の集合をデータとして作成する。研究方針の状態であるため、アブストラクトのうち、実験結果に関する記述をルールベースで省いたアブストラクトを入力として、該当する論文の関連研究の章で引用されている文献のリストを正解とす

る。

(2) 引用要否判定タスク

図には含めていないが、ベース論文のイントロおよび関連研究の章を対象とする。ベース論文中で引用されている文は引用が必要な文と見なし、各文に対して、引用が必要な文かどうかを判定する。

(3) 引用文献割当タスク/引用文・被引用文献ペア適正判定タスク

ベース論文の関連研究の章を対象とする。ベース論文中で引用されている文は正しく引用していると思なし、引用のある各文 (文の長さによって前後の文を含める) に対して、適切に引用されているかを推定する。

(4) 引用文に対応する文献を推薦

ベース論文の関連研究の章を対象とする。ベース論文中で引用されている文は正しく引用していると思なし、引用のある各文 (文の長さによって前後の文を含める) に対して、ベース論文で引用している文献リストから、どの文献を引用すべきかを推定する。

(5) 関連研究の章の自動生成

ベース論文の関連研究の章を対象とする。ベース論文の引用方法は、その論文の観点で適切に引用されていると思なし。関連研究の章 (もしくはそのうちの 1 段落) の抜いた該当論文を入力した時に、その関連研究の章 (もしくは 1 段落) を自動で生成する。

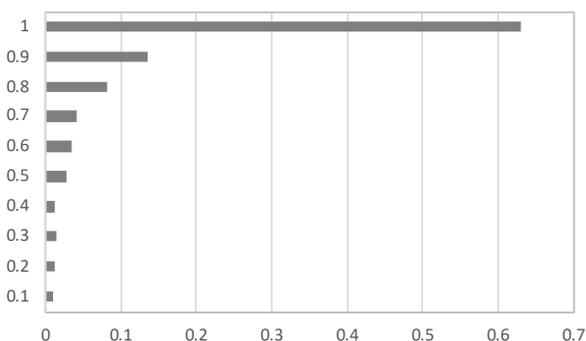


図3 関連研究で引用されている文献のうち本文が取得できる文献の割合（縦軸）と該当する文献の割合（横軸）

5 評価

3節で定義したサブタスクのうち、従来研究で解かれていないタスク(3)のうち(3)-2の引用文献が正しいかどうかを判定するタスクのについて、ベースラインの性能を評価する。

タスクデータは、Leaderbord自動生成を目的としたタスクデータとして公開されているAxCell[8]を元に取得した29,121件の論文をベース論文として構築した。大規模文献DBについては、arXiv API²⁾から可能な限り取得した文献を格納した。なお、関連研究で引用されている文献のうち、全ての文献のtexのソースもしくはPDFが取得できる論文の割合は、全体の6割程度の21,351件であった。図3に、関連研究で引用されている文献のうち、texのソースもしくはPDFが取得できる文献の割合およびそれに該当する文献の割合の内訳を示す。

従来のベースとして、Word2Vec[14]と、近年の大規模汎用言語モデルの元となったBERT[15]を用いて性能を評価する。

BERT[15]では、bert-base-uncasedのモデル³⁾を用い、Document Classification Taskのモデルを以下のフォーマットで学習した。引用文とその引用箇所での引用されている文献のアブストラクトのペアを正例、同じ関連研究の章で引用されている他の文献のアブストラクトをペアとしたときを負例として用いた。学習データは、正例負例ともに18,000ずつ、devとして3,000、テストに3,000用いた。またWord2Vec[14]は、文の平均ベクトルの類似度で判定したときのスコアを比較した。結果を表2に示す。

評価した性能を表2示す。

2) <https://arxiv.org/help/api/>

3) <https://github.com/huggingface/transformers>

[CLS]	引用文	[SEP]	引用文献のアブスト	[SEP]
-------	-----	-------	-----------	-------

図4 BERTに入力したデータのフォーマット

	Precision	Recall	F-measure
Word2Vec	0.541	0.763	0.633
BERT	0.822	0.806	0.814

表2 Word2VecとBERTの転移学習をした場合のタスク(3)-2の精度

約8割程度の正解率であることから、判定不可能ではないタスクであると言える。また、Word2Vecベースと比較すると約20ポイントの伸びがあることから、文脈を踏まえた分散表現により、より高い精度で解ける可能性があると言える。

なお、本タスクの詳細な解法については、文献[16]に記述している。

6 おわりに

本研究では、研究者の研究調査および論文執筆における関連研究の引用および生成に関わる統合的な執筆支援を目的として、関連研究に関わる様々な既存のタスクを統合した新たなデータセット構築方法および5つのタスクを定義した。

定義した5つのタスクのうち、本稿で新たに定義した正しい引用文献かどうかの判定タスクにおいて、近年の言語処理技術のベースとなっているBERTを転移学習した場合の結果を示した。現状の引用文と引用文献のアブストラクトのペアの正しさを判定するモデルの性能は8割程度であるが、論文全体を用いることでさらに精度向上が期待される。

今後は論文全体を入力とした判定モデルの構築手法を検討し、タスクの精度向上に取り組む。また、未着手のその他のタスクについても取り組み、統合的に論文執筆を支援する仕組みの実現を目指す。

参考文献

- [1] Simone Teufel and Marc Moens. Summarizing scientific articles: experiments with relevance and rhetorical status. *Computational linguistics*, Vol. 28, No. 4, pp. 409–445, 2002.
- [2] X. Bai, M. Wang, I. Lee, Z. Yang, X. Kong, and F. Xia. Scientific paper recommendation: A survey. *IEEE Access*, Vol. 7, pp. 9324–9339, 2019.
- [3] Vahed Qazvinian and Dragomir R. Radev. Scientific paper summarization using citation summary networks. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pp. 689–696, Manchester, UK, August 2008. Coling 2008 Organizing Committee.

- [4] Chandra Bhagavatula, Sergey Feldman, Russell Power, and Waleed Ammar. Content-based citation recommendation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pp. 238–251, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.
- [5] Saif Mohammad, Bonnie Dorr, Melissa Egan, Ahmed Hassan, Pradeep Muthukrishnan, Vahed Qazvinian, Dragomir Radev, and David Zajic. Using citations to generate surveys of scientific paradigms. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, NAACL '09, p. 584–592, USA, 2009. Association for Computational Linguistics.
- [6] Michael Färber, Alexander Thiemann, and Adam Jatowt. To cite, or not to cite? detecting citation contexts in text. In Gabriella Pasi, Benjamin Piwowarski, Leif Azzopardi, and Allan Hanbury, editors, *Advances in Information Retrieval*, pp. 598–603, Cham, 2018. Springer International Publishing.
- [7] Michael Färber and Adam Jatowt. Citation recommendation: approaches and datasets. *International Journal on Digital Libraries*, Vol. 21, No. 4, p. 375–405, Aug 2020.
- [8] Marcin Kardas, Piotr Czapla, Pontus Stenetorp, Sebastian Ruder, Sebastian Riedel, Ross Taylor, and Robert Stojnic. Axccl: Automatic extraction of results from machine learning papers. *arXiv preprint arXiv:2004.14356*, 2020.
- [9] Tarek Saier and Michael Färber. unarxive: a large scholarly data set with publications’ full-text, annotated in-text citations, and links to metadata. *aScientometrics*, Vol. 125, No. 3, pp. 3085–3108, 2020.
- [10] Michihiro Yasunaga, Jungo Kasai, Rui Zhang, Alexander R Fabbri, Irene Li, Dan Friedman, and Dragomir R Radev. Scisummnet: A large annotated corpus and content-impact models for scientific paper summarization with citation networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, pp. 7386–7393, 2019.
- [11] T. Chakraborty, N. Modani, R. Narayanam, and S. Nagar. Discern: A diversified citation recommendation system for scientific queries. In *2015 IEEE 31st International Conference on Data Engineering*, pp. 555–566, 2015.
- [12] Pavlos Kefalas Waheed Ahmad Abro Bahadar Ali Zafar Ali, Guilin Qi. A graph-based taxonomy of citation recommendation models. *Artificial Intelligence Review*, Vol. 53, No. 7, pp. 1573–7462, 2020.
- [13] Khalid Haruna, Maizatul Akmar Ismail, Abdullahi Baffa Bichi, Victor Chang, Sutrisna Wibawa, and Tutut Herawan. A citation-based recommender system for scholarly paper recommendation. In *International Conference on Computational Science and Its Applications*, pp. 514–525. Springer, 2018.
- [14] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [15] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [16] 小山康平, 南泰浩, 成松宏美, 堂坂浩二, 田盛大悟, 東中竜一郎, 平博順. 学術論文における関連研究の執筆支援のための被引用論文の推定. 言語処理学会第 26 回年次大会, 2021.