

論文閲覧支援のための引用意図に関する検討

Examination of citation intention for support of browsing research papers

西海 真祥

Masayoshi Nishiumi

岡山大学 太田研究室

Ohta Laboratory, Okayama University

概要 本研究では、論文閲覧支援の1つとして引用箇所の理解を助ける情報を、引用意図に基づいて提示することを目指す。そのため本稿では、Vaswaniらの“Attention Is All You Need”の論文に現れる引用箇所を分析して引用意図クラスを新たに定義する。さらに、画像生成に関する2つの論文の引用箇所の引用意図を手で分類することで、引用意図の事例の分布を確認した。

1 はじめに

学術論文において引用は、先行研究の手法、知見や実験結果を参照するために用いられる。特に被引用文献との関係から、研究の位置付けを示すことが多い。従って、被引用文献の内容を知ることが論文閲覧において重要な意味を持っている。

学術論文の場合、一般的に被引用文献から必要な部分を著者の主観でまとめた形で引用する。そのため、引用箇所を理解するためには被引用文献を必要なだけ読み進めなければならないが、学術論文では一般に引用は多数存在するため、全ての被引用文献を読む時間的なコストは大きい。この問題を解消するために、石井ら [1] は、引用箇所を6つの引用意図クラスに分類し、その分類に従って引用箇所に最も適当な被引用文献中の文章を抽出する手法を提案した。この手法では引用箇所に合致する内容が被引用文献中の一箇所にまとまっていない限り、引用箇所の理解を助けるには不十分であるという問題があった。本研究では、引用箇所の理解をより深められる情報を、引用意図に基づいて提示することを目指す。

本稿では、引用意図によって必要となる情報の傾向が大きく異なる点を考慮し、引用意図を再定義する。さらに、より適切な情報を論文閲覧者に提示するために、引用意図の分類についても検討する。また3論文について引用箇所の引用意図を本稿で定義した引用意図クラスに著者が分類し、事例の分布を調査した。

2 引用意図

2.1 引用意図クラス

石井らは、情報の理解や活用を支援する技術を相互に学ぶプロジェクトである NTCIR-9[2] のタスク参加者の論文を基に引用意図クラスを定義した。そのため、引用意図の定義が NTCIR-9 のようなタスク指向の論文に特化しているという問題があった。そのため、本稿では、著名な国際会議論文の引用箇所を精査し、以降で説明する主クラス、関係性クラスと抽象度クラスの組み合わせで引用意図クラスを再定義する。

2.1.1 主クラス

主クラスは、引用の対象を分類するクラスである。主クラスとして、以下の5つを定義する。

1. Method

研究で採用されたモデルやアルゴリズムを引用の対象とする。

2. Result

研究で得られた主要な結果を引用の対象とする。

3. Basis

研究で利用したパラメータ、考察や議論を引用の対象とする。

4. Data

実験などに用いたデータを引用の対象とする。

5. Other

上記4つのクラスに該当しない内容を引用の対象とする。

2.1.2 関係性クラス

関係性クラスは、引用の理由を表すクラスである。閲覧論文が被引用文献と異なる点、または同じ点を明確にすることは論文閲覧支援に有用であると考え導入する。関係性クラスとして、以下の2つを定義する。

1. Comparison

被引用文献と閲覧論文の内容を対比するために引用されている。

2. Introduction

被引用文献の内容を閲覧論文で導入するために引用されている。

2.1.3 抽象度クラス

抽象度クラスは、引用内容の具体性を表す。引用内容の抽象度が高ければ、被引用文献以外の情報を被引用文章として論文閲覧者に提供可能と考える。抽象度クラスとして、以下の2つを定義する。

1. Instance or Experiment

具体的な手続きや数値が引用対象である。原則、論文閲覧支援として提示する内容は被引用文献の内容から構成する。

2. Concept or Conclusion

抽象度の高い概念やアイデアが引用対象である。論文閲覧支援として提示する内容は被引用文献以外からも選択できる。

2.1.4 引用意図クラスの構成

引用意図クラスは主クラスを関係性クラス、抽象度クラスで拡張して構成する。具体的には Method クラスは関係性クラスと抽象度クラスで拡張し、Basis, Result クラスは抽象度クラスで拡張する。Data, Other クラスは拡張しない。各クラスによって拡張が異なるのは、引用箇所に表示する情報の質への寄与が小さいと考えるクラスは拡張しないためである。

3 引用意図の分析と分類

3.1 分析内容

まず Vaswani ら [3] を用いて本稿で定めた引用意図について分析する。この論文を採用した理由は、刺激的なタイトルと Transformer が示す性能で関心を集めた論文であり、発表当時最先端の論文でありながら閲覧支援が必要な人が多かったと考えたからである。

Vaswani らは [3] の中で 40 本の参考文献を計 76 回引用している。これらの引用箇所を分析し、2 節に示した引用意図クラスに手動で分類した。

3.2 分析結果

Vaswani らの引用箇所の分類結果を表 1 に示す。引用意図毎に引用箇所と引用数をまとめている。主クラスの内、Method と Result クラスに分類された引用箇所が多いことがわかる。これは学術論文の目的から考えても妥当であり、主クラスの分類事例の偏在はやむを得ない。一方、本稿で提案した関係性クラスと抽象度クラスによって大きな主クラスは細分化されている。

表 1 Vaswani ら [3] の引用意図の分類

引用意図		引用箇所	引用数	
Method	Comparison	Instance	[16] [18] [9] [17, 18] [9] [18] [6] [8] [29]	10
		Concept	[13] [7] [21] [32] [2] [37]	6
	Introduction	Instance	[10] [30] [20]	3
		Concept	[5, 2, 35] [11] [1] [38, 2, 9] [9] [9] [38] [31] [3] [33] [9]	15
Result	Experiment	[18] [39] [38] [9] [32] [39] [38] [9] [37] [29] [40] [8] [40] [14] [26] [37] [23] [8]	18	
	Conclusion	[35, 2, 5] [38, 24, 15] [2, 19] [27] [4, 27, 28, 22] [34]	14	
Basis	Instance	[36] [38]	2	
	Concept	[12] [3] [12] [38] [37]	5	
Data		[38] [25] [37]	3	
Other		None	0	
合計			76	

表 2 Goodfellow ら [4] と Razavi ら [5] の引用意図の分類

引用意図		引用数		
		[4]	[5]	
Method	Comparison	Instance	1	1
		Concept	8	20
	Introduction	Instance	2	3
		Concept	16	26
Result	Experiment	3	1	
	Conclusion	5	23	
Basis	Instance	0	1	
	Concept	3	6	
Data		3	2	
Other		3	0	
合計		44	83	

3.3 画像生成に関する 2 論文の引用意図の分類

画像生成に関する 2 つの論文、Goodfellow ら [4] と Razavi ら [5] について引用意図を分類する。[4] で Goodfellow らは 31 本の参考文献を計 44 回、[5] で Razavi らは 40 本の参考文献を計 83 回引用している。

3.4 分類結果

分類結果を表 2 に示す。表 2 では引用意図毎に両論文の引用数をまとめている。抽象度クラスでは Concept クラスが [4] で 84%, [5] で 92% を占めている。特に [4] では Basis×Instance クラスに該当する引用箇所が存在しなかった。

[4] では Other クラスに割り振られた引用箇所が複数あるが、これらはソフトウェア開発者に謝辞を示す際にソフトウェアに対して引用が行われていた箇所である。

4 考察

抽象度クラスは元々、本研究で目指す論文閲覧支援において、情報源の目安や提示する情報の観点に活用できると考えて導入した。特に Method×Concept に

分類された引用箇所はインターネットで検索すると該当件数が多く、これらの情報を論文閲覧支援に活用したい。

関係性クラスは、引用の理由を表すため、論文閲覧支援に有用と考えて導入した。ただし、主クラスや抽象度クラスは引用箇所近傍から判断し易いのに対して、関係性クラスは閲覧論文の全体像を掴まないと判断が難しい箇所があり、他クラスよりも自動分類が難しい可能性がある。

5 まとめ

本稿では、論文閲覧支援のために論文中の引用箇所の引用意図クラスを定義し、3つの論文に対して引用意図分類を手動で行うことで、論文閲覧支援のための引用意図クラスの妥当性を分析した。まだ分析事例が少ないため、より多くの事例について引用意図クラスの妥当性について検証する必要がある。また、定めた引用意図に応じて、論文閲覧支援のため引用箇所に提示する情報についても検討する必要がある。

参考文献

- [1] 石井仁子 他, “引用意図を利用した学術論文閲覧支援のための適切な被引用箇所の特定,” 第7回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2015), F3-5, 2015.
- [2] “NTCIR-9 (the 9th NII Test Collection for IR Systems),” <http://research.nii.ac.jp/ntcir/ntcir-9>
- [3] Ashish Vaswani, *et al.*, “Attention Is All You Need,” *Advances in Neural Information Processing Systems* 30 (NIPS2017), pages 6000–6010, 2017.
- [4] Ian Goodfellow, *et al.*, “Generative adversarial nets,” *Advances in neural information processing systems* 27 (NIPS2014), pages 2672–2680, 2014.
- [5] Ali Razavi, *et al.*, “Generating Diverse High-Fidelity Images with VQ-VAE-2,” arXiv preprint arXiv:1906.00446, 2019.