

被引用文特定のための ChatGPT の利用に関する検討

Examination of Using ChatGPT for Identifying Cited Sentences

佐川 俊輔

Syunsuke Sagawa

岡山大学 太田研究室

Ohta Laboratory, Okayama University

概要 学術論文では通常多数の参考文献が引用されているため、被引用論文中の被引用箇所を特定することは重要である。本稿では、被引用箇所にふさわしい一文を被引用文として、学術論文中の引用文から被引用文を特定する方法を検討する。西海ら [2] は、文の類似度と Extractive Question Answering(EQA) の2つのタスクを用いた方法を提案し、被引用文を特定した。本稿では、ChatGPT を用いて被引用文を特定する実験を行い、その結果を西海らの手法と比較する。

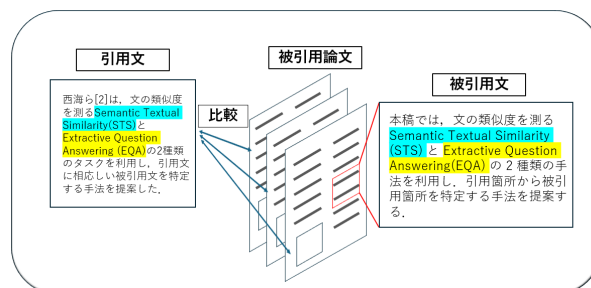


図 1: 被引用文特定の概要

1 はじめに

学術論文では通常多数の参考文献が引用されており、引用内容を把握することが論文を理解する上で重要となる。しかし、全ての参考文献を参照し、その内容を理解するには大きな労力と時間がかかる。そこで、適切な被引用文を特定し、それを論文閲覧者に提供することができればその負担は小さくなる。

本稿では、被引用文特定について説明し、先行研究を紹介する。また、引用文から被引用文を特定するために、ChatGPT を用いた手法を検討する。また、ChatGPT に論文とその被引用論文を与え、被引用文を特定する実験を行ったので報告する。

本稿では、2 節で被引用文の特定手法の先行研究として西海ら [2] の手法を紹介し、3 節で ChatGPT による被引用文の特定方法を説明する。4 節で ChatGPT による被引用文特定の実験について考察し、5 節でまとめる。

2 被引用文の特定手法

2.1 被引用文特定

被引用文の特定とは、CL-SciSumm Shared Task[1]において、引用箇所引用した内容を最も正確に反映している文を被引用論文から特定することと定義されている。被引用文特定の概要を図 1 に示す。引用文と被引用論文を比較することで、被引用文を特定する。The 2020 Conference on Empirical Methods in Natural Language Processing(EMNLP 2020) の ASD ワークショップの CL-SciSumm Shared Task[1] では、学術論文を要約するタスクがあり、そのサブタスクとして、引用文から被引用文を特定するタスクが存在する。

2.2 STS と EQA による被引用文特定

西海ら [2] は、文の類似度を測る Semantic Textual Similarity (STS) と Extractive Question Answering

(EQA) の 2 種類のタスクを利用し、引用文にふさわしい被引用文を特定する手法を提案した。以下に STS と EQA を説明する。

- Semantic Textual Similarity (STS)

STS は、2 つのテキスト間の意味的な類似性を測定するタスクである。具体的には、異なるテキスト同士がどれだけ意味的に似ているかを数値で表す。この数値は通常 0 から 5 の範囲で表され、0 が全く意味が似ていない、5 が完全に同じ意味を持つことを表す。

- Extractive Question Answering (EQA)

EQA は、与えられたテキスト (文書や段落など) から直接回答を抽出するタスクである。具体的には、ユーザーが質問を入力すると、システムはその質問に対する回答をテキストの中からそのまま取り出して返答する。このタスクでは、回答は常に入力テキスト内に含まれている。

西海ら [2] は、被引用文の特定では Sentence-BERT[3] を使用し、結果を出力している。Sentence-BERT は BERT を改良したもので、BERT と同等の精度を維持しつつ、より効率的に動作するモデルである。

3 ChatGPT による被引用文の特定

3.1 ChatGPT

ChatGPT^{*1}は、OpenAI が公開している人工知能チャットボットであり、大規模言語モデルに基づいて構築されている。質問や指示形式のプロンプトを与えることによって、それに対応した回答が生成される。

*1 <https://openai.com/>

表 1: 被引用文特定のためのプロンプト

入力	
(1)	これからある論文をまとめたデータを与えます。各文にはそれぞれ S sid と ssid が登録されています。データは複数回に分けて与えるので、全文記録し、次の命令が来るまで待機してください。
(2)	<PAPER><S sid= “0” > A Maximum-Entropy-Inspired Parser *</S><ABSTRACT><S sid=“1” ssid= “1” >We present a new parser for
(3)	次にこの論文を引用した文を与えるので、論文のどの文を引用したか予測し、予測結果の上位 2 文とその文に登録されている S sid を提示してください。なお、引用文は予測後に記録から削除してください。以下が引用文になります。 We then use Charniak’s parser (Charniak, 2000) to generate the most likely parse tree for each English target sentence in the training corpus.

3.2 ChatGPT による被引用文の特定方法

本稿では、ChatGPT を利用して被引用文を特定する。

表 1 に使用したプロンプトを示す。入力 (1) で被引用論文と sid, ssid の 2 つ ID について説明する。sid は、その文が論文全体の何文目に該当するか、ssid は Abstract や Conclusion などの節の中で何文目に該当するか表すものである。入力 (2) で被引用論文のデータを与える。入力 (3) で被引用論文から被引用文を予測するように命令し、引用文を与える。

4 ChatGPT による被引用文の特定実験

4.1 実験概要

ChatGPT-3.5 を用いて被引用文の特定実験をした。実験では、被引用論文と引用文のデータのみを与えて、被引用文と予測される上位 2 文を特定した。また、このデータは、西海ら [2] が実験に使用したデータの一部であり、被引用論文が 14 件、引用文が 226 文ある。評価指標は、以下の式に示す正解率を使用する。

$$\text{正解率} = \frac{\text{被引用文と予測した文のうち実際に被引用文であった数}}{\text{引用文の数}}$$

4.2 実験結果

ChatGPT-3.5 と西海らの手法の実験結果を表 2 にまとめる。ChatGPT-3.5(1st) が予測結果の 1 位のものであり、ChatGPT-3.5(2nd) が予測結果の 2 位のものである。それぞれの正解数は 1 位のみ正解、2 位のみ正解

表 2: 実験結果

モデル	引用文数	正解	不正解	正解率
ChatGPT-3.5(1st)	226	18	208	0.079646
ChatGPT-3.5(2nd)	226	6	220	0.026549
西海ら [2]	226	87	139	0.384956

した場合の数である。正解率は、ChatGPT-3.5(1st) と西海らのモデルで、約 30 ポイントとかなり差が出た。実験で ChatGPT-3.5 に与えたデータは被引用論文と、引用文のみであり、正解例と引用論文を提示しなかった。そこで、被引用論文と引用文に加えて、閲覧論文のデータも ChatGPT-3.5 に与えて被引用文を予測させる追加実験をした。被引用論文 2 件、引用文 33 件で実験したところ、閲覧論文を与えなかった場合と精度に違いはなかった。

また、この実験では、ChatGPT-3.5 に引用文を 1 件ずつ与えて被引用文を特定したが、1 回あたりの実行時間は非常に短く、入力するとほぼ同時に結果が出力され始めた。実験では引用文 226 件の全ての結果を記録するのに約 4 時間ほどかかった。

5 まとめ

本稿では、学術論文における被引用文の特定手法について説明し、ChatGPT による被引用文特定実験について述べた。実験の結果、ChatGPT-3.5 と西海らのモデルでは、西海らのほうが正解率が約 30 ポイント高く、精度の向上を図る必要があると分かった。

今後の課題として、ChatGPT による被引用文特定では、プロンプトや与えるデータなどを変更して正解率がどれだけ変化するか確認し、より精度を向上させる方法を検討したい。

参考文献

- [1] Chandrasekaran, M. K. *et al.*, “Overview and Insights from the Shared Tasks at Scholarly Document Processing 2020: CL-SciSumm, LaySumm and LongSumm,” *In Proceedings of the First Workshop on Scholarly Document Processing, Association for Computational Linguistics*, pp. 214–224 (online), 2020.
- [2] 西海真祥, 金澤輝一, 上野史, 太田学, “文の類似度と Extractive QA による被引用文特定の一手法,” DEIM Forum 2023, 1-b-4-5. 2023.
- [3] Reimers, N. *et al.*, “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks,” *In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, (online), 2019.