

脱文脈化に基づく RAG のマルチホップ QA への適用

尾崎花奈 今一修 平山淳一

株式会社日立製作所 研究開発グループ

{kana.ozaki.dj, osamu.imaichi.xc, junichi.hirayama.qq}@hitachi.com

概要

本研究では、マルチホップ QA タスクに対し、文書の脱文脈化と文脈補完質問文の事前生成を組み合わせた新たな RAG フレームワーク「DeconRAG」を提案する。3種類のマルチホップ QA データセットで実験を行い、DeconRAG は従来の NaiveRAG や質問分解のみを用いた手法よりも高い精度を示した。また、多段推論型 RAG 手法である ChainRAG と比較しても、DeconRAG は2種類のデータセットで上回りつつ、推論効率にも優れることが確認された。

1 はじめに

マルチホップ QA タスクは、複数の断片的な知識や文脈から情報を統合し、複雑な質問に答える能力が求められる重要なタスクである。近年、このマルチホップ QA タスクに対して広く利用されているのが Retrieval Augmented Generation (RAG) [1] である。RAG は、大規模言語モデル (LLM) の推論・生成能力と外部知識検索を組み合わせることで、高い性能を発揮している。

従来の RAG システムでは、質問分解や多段階ステップ推論など、LLM の推論能力を活用したさまざまな拡張が提案されてきた。これらの手法では、サブ質問生成・検索・回答生成を多段階にわたって繰り返し行うことで、マルチホップ QA への適応が図られている [2][3][4]。

しかし、上記の先行研究の多くは推論過程に着目している一方で、検索によって抽出されるパッセージの文脈依存性の問題には十分に対応できていない。また、逐次的なサブ質問生成や推論ステップを複数回行う方式では、推論ごとに LLM を呼び出す必要があり、即時応答性が求められる QA タスクにおいては計算時間やコストが大きな課題となる。

これらの課題を解決するために、本研究では、検索対象となる文書を脱文脈化 (Decontextualization) [5] し、元の文との差分から生成した質問を検索対象と

するアプローチ「DeconRAG」を提案する。脱文脈化とは、各文がそれ自体で意味をなす独立した文となるよう変換することである。脱文脈化された文書を利用することで、文同士の依存関係に左右されることなく、LLM が必要な知識を正確かつ効率的に抽出できることが期待される。さらに、脱文脈化された文と元の文との差分から質問を生成することで、マルチホップ QA において質問文を分解したサブ質問に対する検索精度の向上も見込める。本論文では、3つのマルチホップ QA データセットを用いて提案手法の評価を行い、回答精度と効率性の両面からその有効性を検証する。

2 関連研究

2.1 Retrieval Augmented Generation

Retrieval Augmented Generation (RAG) [1] は、大規模言語モデル (LLM) の生成能力と外部知識検索を統合することで、高度な質問応答や知識推論を実現する枠組みである。近年は、質問文拡張 [6, 7, 8]、LLM による能動的な検索制御 [9, 10]、および文書構造のグラフ化を取り入れた GraphRAG [11] など、多様な改良技術が提案されている。

特にマルチホップ QA タスクに対しては、質問文を複数のサブ質問に分解し、検索・推論を段階的に行うことで、LLM の推論能力を最大限活用して精度を向上させる手法が主流となっている [2, 3, 4]。ChainRAG [2] では、質問文分解時に重要なエンティティが失われる「lost-in-retrieval」現象を防ぐために、サブ質問の回答から次のサブ質問へと欠落エンティティを補完する質問文の書き換えを導入したことで精度が向上したと報告している。

2.2 脱文脈化

脱文脈化 (Decontextualization) は、文や断片的なパッセージをそれ自体で解釈可能かつ文脈依存性のない「独立文」として再生成する。Choi ら [5] はこの課題

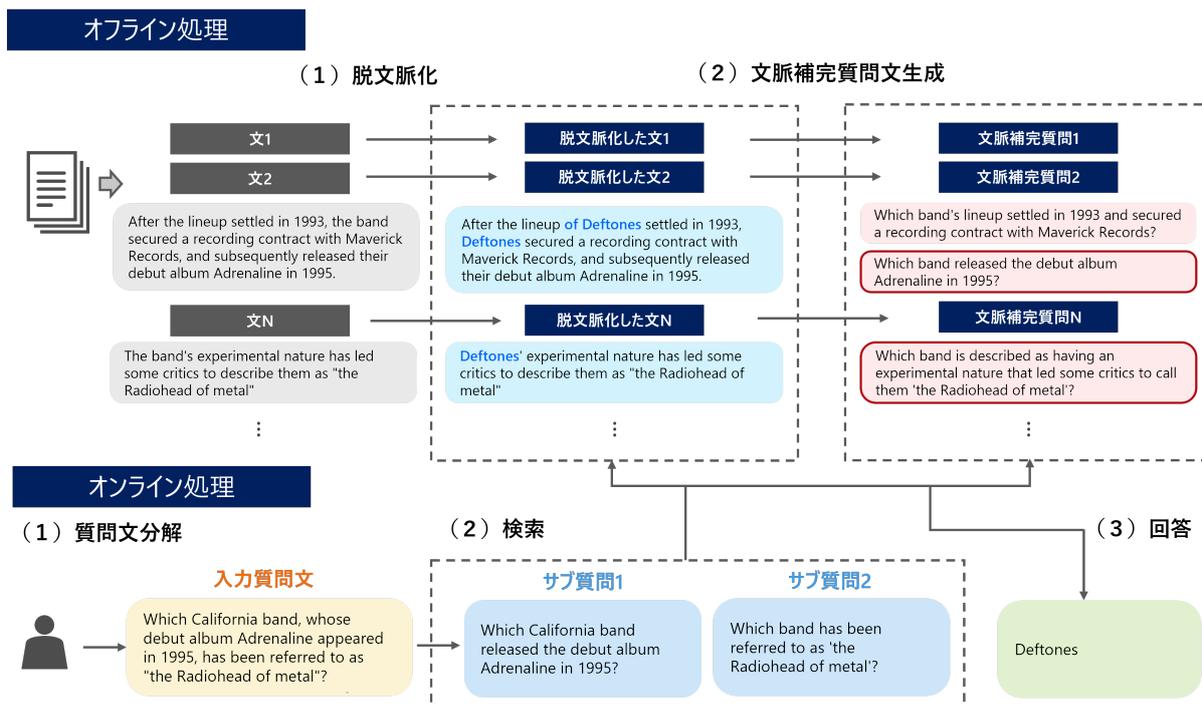


図 1 提案手法「DeconRAG」の概要図. オフライン処理にて脱文脈化により事前に生成しておいた質問を, オンライン処理にて入力質問文のサブ質問を用いて検索することでマルチホップ QA における検索を多段推論なしに実現する.

を定式化し, 名詞句置換, 修飾語追加などの編集タイプを分類したデータセットを構築した. 加えて, 共参照解析や生成モデルの活用による自動化と評価も行われ, NLP 応用への有効性が示されている.

近年は LLM による Few-shot decontextualization [12] や, 科学文書のスニペット生成 [13] など手法自体の発展が進む一方, LLM の生成文を脱文脈化された文に分解してファクトチェックを行う方法 [14, 15] が提案されるなど. 脱文脈化は応用的な NLP タスクに活用され始めている.

3 提案手法

本稿では, マルチホップ QA タスクにおける文脈依存性の課題を解決するため, 文書の各文を脱文脈化し, さらに元文との差分から文脈を補完する文脈補完質問文を生成する「DeconRAG」を提案する. 図 1 に本手法の全体構成を示す.

3.1 オフライン処理

まず, 文書内の各文に対して LLM を用いて脱文脈化処理を施す. 具体的なプロンプトは付録 A.2 の図 2 に示す. これは, 各文が周辺文に依存せずそれぞれで意味が通る独立した文となるように書き換えるプロセスであり, LLM が文単体を入力としても意味を正確に把握しやすくなる.

さらに, 脱文脈化した文と元文との差分情報から, LLM を用いて文脈補完質問文を生成する. 推論時に必要となる文脈依存的な情報をあらかじめ質問として切り出しておくことで, 検索時に文脈に依存した不足情報を明示的に補完できるようになる. 具体的なプロンプトは付録 A.2 の図 3 に示す.

3.2 オンライン処理

推論段階では, ユーザからの質問を受け取り, LLM を用いて単一の答えが定まるサブ質問に分割する. 質問分割は ChainRAG [2] に準じ, 必要な場合のみ分解し, 不要な場合はそのままとする 2 段階のステップを採用する.

検索は, 各サブ質問ごとに, 以下の順序で N 文になるまで対応する脱文脈化文を収集する. まず, 埋め込みベクトルの類似度が T 以上であり, かつ質問文の疑問詞が一致する文脈補完質問文に対応する脱文脈化文を優先的に追加する. 次に, 疑問詞は一致しなくても類似度が閾値 T 以上となる文脈補完質問文に対応する脱文脈化文を, 類似度が高い順に追加する. 最後に, 文書全体の脱文脈化文を検索対象とし, 類似度の高い順に追加する.

このような段階的な検索で収集した各サブ質問に関連する脱文脈化文 N 文とユーザ質問文を LLM の入力として, 最終回答の生成を行う. すなわち,

| 手法 | QD | MuSiQue | | 2Wiki | | HotpotQA | |
|-----------------------|----|--------------|--------------|--------------|--------------|--------------|--------------|
| | | F1 | EM | F1 | EM | F1 | EM |
| NaiveRAG | - | 28.90 | 20.00 | 37.71 | 33.00 | 44.36 | 31.50 |
| NaiveRAG (QD) | ○ | 26.25 | 20.50 | 37.05 | 32.00 | 43.67 | 31.00 |
| ChainRAG | ○ | 38.74 | 31.50 | <u>49.43</u> | <u>43.00</u> | 46.60 | 36.50 |
| DeconRAG wo/質問生成 | - | 30.04 | 23.00 | 35.74 | 30.50 | <u>54.87</u> | <u>42.50</u> |
| DeconRAG wo/質問生成 (QD) | ○ | 27.10 | 21.50 | 36.81 | 32.00 | 54.32 | 40.50 |
| DeconRAG | ○ | <u>37.93</u> | <u>29.00</u> | 51.40 | 44.50 | 59.70 | 46.50 |

表1 評価結果 (%). QDは質問文の分解 (Question Decomposition), 太字は最高評価値, 下線は次点の評価値を示す.

LLMによる回答生成処理は, ユーザ質問に対する最終回答時の1度だけであり, ChainRAGのように各サブ質問ごとにLLMを繰り返し呼び出すことはない. この設計により, 推論過程でのLLM呼び出し回数を大幅に削減することができる.

4 実験

4.1 実験設定

マルチホップQAデータセットは, MuSiQue [16], 2WikiMultiHopQA (2Wiki) [17], HotpotQA [18]を用いた. ChainRAG [2]に倣い, 3つのデータセットから各200件のQAをランダムに抽出した. データセットの詳細はAppendix A.1に示す.

評価指標には, F1スコアとExact Match (EM)スコアを用い, 以下の3手法との比較を行った.

NaiveRAG 基本的なRAG. 入力質問文で全文を検索し, ヒットした文をコンテキストとしてLLMの入力とする.

NaiveRAG (QD) 質問文を分解 (Question Decomposition) し, 分解した各サブ質問にヒットした文をコンテキストとしてLLMの入力とする.

ChainRAG [2] オフライン処理では, 各文を近接性やエンティティの共通性を利用してグラフ化をする. オンライン処理では, 先行するサブ質問の回答を次のサブ質問の書き換えに用いる. 各サブ質問でヒットした文とグラフ上の近傍の文をコンテキストとしてLLMの入力とする.

また, 提案手法のバリエーションとして, 文書内の各文を脱文脈化した文のみを検索対象とし, 文脈補完質問文の生成は行わない「DeconRAG wo/質問生成」についても, 入力質問文を分解する場合と分解しない場合の両方で実験を行った.

提案手法, 比較手法ともに埋め込みモデルは text-

| 手法 | MuSiQue | 2Wiki | HotpotQA |
|----------|---------|-------|----------|
| ChainRAG | 10.08 | 9.61 | 8.89 |
| DeconRAG | 2.93 | 2.98 | 2.81 |

表2 各手法におけるLLM呼び出しの平均回数.

embedding-small-v3¹⁾を, LLMはGPT-4.1-mini²⁾を用いた. 文脈補完質問文の検索時のコサイン類似度の下限 T は0.8とした. また, すべての手法で各サブ質問に対して検索する文の数 N は3とした. なお, ChainRAGでは, 文グラフを検索する際に用いるseedとなる文の数を1に設定した.

4.2 結果

結果を表1に示す. 全体として, DeconRAGは従来手法と比較して優れた結果を示した. まず, NaiveRAGやNaiveRAG (QD)と比較して, DeconRAGは全データセットにおいてF1スコア, EMスコア共に高スコアを記録した. 特に2Wikiでは13%, HotpotQAでは15%とF1スコアが大きく向上した.

ChainRAGとの比較では, MuSiQueではChainRAGが最も高いF1スコアとなっているものの, 2WikiとHotpotQAではDeconRAGが上回った. HotpotQAでのF1スコアはDeconRAGがChainRAGを約13%上回っている. また, DeconRAGは質問分解と最終回答のみLLMを利用するのに対し, ChainRAGは推論過程でLLMを複数回呼び出す. 表2に示す通り, DeconRAGの平均呼び出し回数は2.91回であるのに対し, ChainRAGでは9.53回と大幅に多い. このことから, DeconRAGは計算コストを大幅に抑えながら, ChainRAGに匹敵する, あるいはそれ以上の性能を実現できているといえる.

さらに, DeconRAG wo/質問生成とDeconRAGを

1) <https://platform.openai.com/docs/guides/embeddings>

2) <https://platform.openai.com/docs/models/gpt-4.1-mini>

| 項目 | 内容 |
|--------------------|--|
| 入力質問文 | For which band, was the female member of Fitz and The Tantrums, the front woman for ten years? |
| サブ質問 1 | Who is the female member of Fitz and The Tantrums? |
| ヒットした 文脈補完質問文 1 | Who became a co-lead singer of the indie pop and neo-soul band Fitz and the Tantrums in 2008? |
| サブ質問 2 | For which band was this female member the front woman for ten years? |
| ヒットした 文脈補完質問文 2 | What is the full name of the person who toured with the band for ten years before taking two years off from music? |
| 正解回答 | The Rebirth |
| ChainRAG の回答 | Angel Nation |
| DeconRAG の回答 | The Rebirth |

表 3 DeconRAG で文脈補完質問文にヒットして正解を導出できた例。

比較すると、全データセットでスコアが向上し、文脈補完質問文の生成が有効であることが分かる。特に MuSiQue と 2Wiki においては、NaiveRAG と比較して検索対象を脱文脈化文にするだけでは変化が無かったが、文脈補完質問文を生成した DeconRAG では大幅にスコアが改善された。これは、文書の脱文脈化だけではマルチホップ質問への対応に限界があり、文脈補完質問文の生成が必要な情報の抽出・統合に寄与していることを示している。加えて、NaiveRAG や DeconRAG wo/質問生成では、単に質問分解を行っただけではスコアがほとんど変化しない、あるいはかえって下がる傾向にあり、単純な質問分解のみでは文脈依存性や情報欠落に対応できないことが示唆される。

4.3 事例分析

表 3 は、ChainRAG では正解を導き出せなかったのに対し、DeconRAG は正しい回答に到達できた具体的な事例を示している。

このケースでは、元の入力質問文の答えを得るために「Fitz and The Tantrums」の女性メンバーが誰なのかを特定し(サブ質問 1)、次にその人物が 10 年間リーダーを務めたバンド名を明らかにする(サブ質問 2)というマルチホップな推論が求められる。ChainRAG では、サブ質問の回答を次のサブ質問の書き換えに利用する設計となっているが、本例ではサブ質問の書き換えに失敗し、最終的に誤った回答を出力していた。

一方、DeconRAG では、まず「Who became a co-lead singer of the indie pop and neo-soul band Fitz and the Tantrums in 2008?」という文脈補完質問によって女

性メンバー (Noelle Scaggs) の情報を正確に取得し、さらに「What is the full name of the person who toured with the band for ten years before taking two years off from music?」という補完質問を通じて、10 年間フロントウーマンを務めたバンド「The Rebirth」の情報を的確に抽出した。これにより、最終的な答えとして正解「The Rebirth」に到達している。

このように、DeconRAG は脱文脈化と文脈補完質問を組み合わせたサブ質問検索によって、従来手法では苦手としていた文脈依存性や情報分散といった課題に対しても、必要な文脈情報を補足して正解に導くことができる。本事例は、文脈補完質問生成と検索フレームワークの有効性により、複数文を横断する情報統合を多段の推論なしで実現可能であることを示している。

5 おわりに

本論文では、文脈依存性の高いマルチホップ QA に対して、文書の脱文脈化と文脈補完質問文の事前生成を組み合わせた新しい RAG フレームワーク「DeconRAG」を提案した。実験の結果、DeconRAG は NaiveRAG や質問分解のみの手法を大きく上回り、さらに多段推論型 RAG 手法である ChainRAG と比較しても、2つのデータセットで精度・推論効率ともに優れることが確認された。

今後は、文脈補完質問生成手法の最適化、既存の推論アルゴリズムとの統合などを通じて、さらなる性能向上や実応用への展開をめざす。

参考文献

- [1] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Kütler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks. In **Proceedings of the 34th International Conference on Neural Information Processing Systems**, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc.
- [2] Rongzhi Zhu, Xiangyu Liu, Zequn Sun, Yiwei Wang, and Wei Hu. Mitigating lost-in-retrieval problems in retrieval augmented multi-hop question answering. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, **Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 22362–22375, Vienna, Austria, July 2025. Association for Computational Linguistics.
- [3] Paul J. L. Ammann, Jonas Golde, and Alan Akbik. Question decomposition for retrieval-augmented generation. In Jin Zhao, Mingyang Wang, and Zhu Liu, editors, **Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)**, pp. 497–507, Vienna, Austria, July 2025. Association for Computational Linguistics.
- [4] 許韜, 村山友里, 和泉潔. LLM を用いたマルチホップ質問応答における信頼性の高い推論ステップ選出による正解率向上. 人工知能学会論文誌, Vol. 40, No. 5, pp. A-P21_1–13, 2025.
- [5] Eunsol Choi, Jennimaria Palomaki, Matthew Lamm, Tom Kwiatkowski, Dipanjan Das, and Michael Collins. Decontextualization: Making sentences stand-alone. **Transactions of the Association for Computational Linguistics**, Vol. 9, pp. 447–461, 2021.
- [6] Xinbei Ma, Yeyun Gong, Pengcheng He, Hai Zhao, and Nan Duan. Query rewriting in retrieval-augmented large language models. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, **Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing**, pp. 5303–5315, Singapore, December 2023. Association for Computational Linguistics.
- [7] Luyu Gao, Xueguang Ma, Jimmy Lin, and Jamie Callan. Precise zero-shot dense retrieval without relevance labels. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1762–1777, Toronto, Canada, July 2023. Association for Computational Linguistics.
- [8] Huaixiu Steven Zheng, Swaroop Mishra, Xinyun Chen, Heng-Tze Cheng, Ed H. Chi, Quoc V Le, and Denny Zhou. Take a step back: Evoking reasoning via abstraction in large language models. In **The Twelfth International Conference on Learning Representations**. The Twelfth International Conference on Learning Representations, 2024.
- [9] Zhengbao Jiang, Frank Xu, Luyu Gao, Zhiqing Sun, Qian Liu, Jane Dwivedi-Yu, Yiming Yang, Jamie Callan, and Graham Neubig. Active retrieval augmented generation. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, **Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing**, pp. 7969–7992, Singapore, December 2023. Association for Computational Linguistics.
- [10] Weihang Su, Qingyao Ai, Jingtao Zhan, Qian Dong, and Yiqun Liu. Dynamic and parametric retrieval-augmented generation. In **Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval**, SIGIR '25, p. 4118–4121, New York, NY, USA, 2025. Association for Computing Machinery.
- [11] Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, Dasha Metropolitan, Robert Osazuwa Ness, and Jonathan Larson. From local to global: A graph rag approach to query-focused summarization. **CoRR**, 2024.
- [12] Benjamin Kane and Lenhart K. Schubert. Get the gist? using large language models for few-shot decontextualization. **ArXiv**, Vol. abs/2310.06254, , 2023.
- [13] Benjamin Newman, Luca Soldaini, Raymond Fok, Arman Cohan, and Kyle Lo. A question answering framework for decontextualizing user-facing snippets from scientific documents. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, **Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing**, pp. 3194–3212, Singapore, December 2023. Association for Computational Linguistics.
- [14] Anisha Gunjal and Greg Durrett. Molecular facts: Desiderata for decontextualization in LLM fact verification. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, **Findings of the Association for Computational Linguistics: EMNLP 2024**, pp. 3751–3768, Miami, Florida, USA, November 2024. Association for Computational Linguistics.
- [15] Zhenyun Deng, Michael Schlichtkrull, and Andreas Vlachos. Document-level claim extraction and decontextualisation for fact-checking. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, **Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 11943–11954, Bangkok, Thailand, August 2024. Association for Computational Linguistics.
- [16] Harsh Trivedi, Niranjan Balasubramanian, Tushar Khot, and Ashish Sabharwal. Musique: Multihop questions via single-hop question composition. **Transactions of the Association for Computational Linguistics**, Vol. 10, pp. 539–554, 05 2022.
- [17] Xanh Ho, Anh-Khoa Duong Nguyen, Saku Sugawara, and Akiko Aizawa. Constructing a multi-hop QA dataset for comprehensive evaluation of reasoning steps. In Donia Scott, Nuria Bel, and Chengqing Zong, editors, **Proceedings of the 28th International Conference on Computational Linguistics**, pp. 6609–6625, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics.
- [18] Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William Cohen, Ruslan Salakhutdinov, and Christopher D. Manning. HotpotQA: A dataset for diverse, explainable multi-hop question answering. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii, editors, **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing**, pp. 2369–2380, Brussels, Belgium, October-November 2018. Association for Computational Linguistics.

A 付録

A.1 データセット

データセットの詳細を表 4 に示す。ChainRAG [2] に倣い、各オリジナルデータセットからランダムに各 200 件の QA を抽出したものを利用し、文分割には spaCy ライブラリの `en_core_web_sm` モデルを用いた。

| データセット | MuSiQue | 2Wiki | HotpotQA |
|--------|---------|-------|----------|
| サンプル数 | 200 | 200 | 200 |
| 平均文数 | 489 | 224 | 410 |
| 平均単語数 | 11,231 | 4,887 | 9,162 |

表 4 評価データセットの詳細

A.2 プロンプト

入力文書の脱文脈化に用いたプロンプトを図 2 に、文脈補完質問文の生成に用いたプロンプトを図 3 に示す。

脱文脈化プロンプト

Task: Decontextualize target sentence from the given document context.

Definition (Decontextualization):

Given a sentence-context pair (s, c), a sentence `s_2` is a valid decontextualization of `s` if:

- (1) the sentence `s_2` is interpretable in the empty context
- (2) The truth-conditional meaning of `s_2` in the empty context is the same as the truth-conditional meaning of `s` in context `c`.

Instructions and Rules:

- Process sentences individually, in the given order of the input.
- Use information available anywhere in the document to resolve pronouns, anaphora, and elided subjects/objects into explicit entity mentions as needed.
- Prefer keeping proper noun abbreviations, optionally adding the long form at first mention in parentheses when the long form appears in the document. Do not invent expansions not present in the document.
- Keep voice and style as much as possible; do not introduce external facts. If a sentence is already self-contained, keep it unchanged.

Output Format:

Output plain text only; no extra commentary.

Context:

{full_document}

Target sentence:

{target_sentence}

Decontextualized sentence:

図 2 脱文脈化に用いたプロンプト。

文脈補完質問文生成プロンプト

Task:

Generate question and answer pairs based on the differences in informational content between the original sentence and the decontextualized sentence.

Definition (Decontextualization):

(脱文脈化プロンプトの定義と同一)

Instructions and Rules:

- Compare the original sentence ('original_sentence') with the decontextualized sentence ('decontextualized_sentence').
- If there are differences in informational content or meaning between the two sentences, generate questions whose answers are precisely these differences.
- For each distinct piece of missing or altered information, generate a separate question-answer pair.
- If there is no difference in informational content, indicate that no such pairs exist.
- You must not mention or refer to "the original sentence," "the decontextualized sentence," or to either sentence or to the process of comparison in any way in the questions or answers.
- All questions must be decontextualized and explicit enough so that their meaning is clear even without seeing the original or decontextualized text.

For example, questions should contain enough context about what is being asked (such as mentioning specific people, events, objects, or actions relevant to the difference), so that someone reading only the question can fully understand what is being asked and what kind of answer is expected.

- Avoid vague or generic questions (e.g., "What is mentioned?", "How are the children listed?") unless the context is perfectly clear from the question itself.

Example (bad): "How are the children listed?"

Example (good): "Which children are listed as members of the Smith family?"

Example (bad): "Who directed the film mentioned in the text?"

Example (good): "Who directed the 1994 film 'Pulp Fiction'?"

Example (bad): "How many people are mentioned?"

Example (good): "How many children does John have before his remarriage?"

- The output must strictly follow the JSON format specified below.

Output Format:

(JSON フォーマットを指定)

Input Example:

- Original text:

{original_sentence}

- Decontextualized text:

{decontextualized_sentence}

Output:

(Provide your JSON-formatted answer here.)

図 3 文脈補完質問文の生成に用いたプロンプト。