

検索クエリログを用いない自然な質問のマイニングの検討

大村 和正¹ 石原 祥太郎¹

¹ 株式会社日本経済新聞社

{kazumasa.omura,shotaro.ishihara}@nex.nikkei.com

概要

情報欲求に起因して自然発生する質問（自然な質問）のマイニングには様々な応用がある。しかし、素朴なマイニング源として考えられる検索クエリログは一般利用可能でない場合が多い。本研究では、検索クエリログの代わりにテキスト生成モデルを用いた自然な質問のマイニング手法を提案する。具体的にはまず、情報欲求の対象を指定し、これに関する質問の観点をテキスト生成モデルに予測させる。次に、大規模言語モデルを用いて情報欲求の対象と予測された質問の観点を表す2つのキーワードから自然言語の質問文を生成する。実験の結果、検索クエリログを用いない自然な質問のマイニングができる可能性を示唆した。

1 はじめに

「引っ越しに伴う必要な手続きを教えてください」といった、情報欲求に起因して自然発生する質問を Natural Questions（自然な質問）[1, 2] と呼ぶ。自然な質問のマイニングには様々な応用がある。例えば、大規模言語モデル向けの指示学習（質問応答）データセットの人手構築に対して、質問の作成コスト軽減や多様化が期待できる。また、商品・サービスに対する自然な質問が収集できれば、需要分析に有用である。このように、自然な質問のマイニングには一定の需要がある。

ウェブが普及した現代では、多くのユーザが情報を欲し、質問をする際に Google 検索といった検索エンジンを用いる。検索エンジンには通常、1語目に情報欲求の対象を、2語目以降に AND 検索の形で質問の観点を入力する（図 1）。ここで、既存研究 [3] の言葉を借りて、1語目をクエリフォーカス、2語目以降を情報要求観点またはサジェストと呼ぶ。本研究では、自然な質問をクエリフォーカスと情報要求観点を表すキーワード列で相当程度表現できるとみなす。

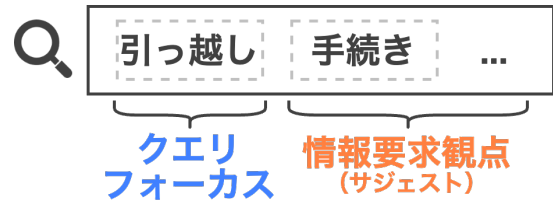


図1 検索エンジンのクエリ例と用語の対応。

自然な質問をマイニングする素朴なアプローチとして、検索エンジンのクエリログを利用すること [1, 2] が挙げられる。しかし、検索クエリログは一般利用可能でなく¹⁾、自前の検索エンジンを持たない個人・企業が自然な質問をマイニングする手法は自明でない。また、コーパスから疑問表現を手がかりに質問文を抽出する [5] というアプローチも考えられるが、ノイズに対処し、内容を分析するために抽出された質問文から情報欲求の対象や質問の観点を解析する必要がある。

その他の取り組みに目を向けると、自然な質問のマイニングの関連タスクとして、与えられたテキスト・画像の内容に対する質問文を生成する Question Generation（質問生成タスク）[6, 7] がある。しかし、質問生成タスクで盛んに取り組まれているのは答えが短文となる質問文の生成であり [8]、多様な質問文の収集は主目的でない。

本研究では、検索クエリログが利用可能でない状況を考え、テキスト生成モデルを用いた自然な質問のマイニング手法を提案する。具体的にはまず、クエリフォーカスを指定し、その情報要求観点がマスクされたテンプレート文を生成する。次に、テキスト生成モデル T5 [9] を用いてマスクされたスパンを予測させることで、情報要求観点の候補を獲得する。最後に、大規模言語モデル (LLM) を用いて指定したクエリフォーカスと予測された情報要求観点から自然言語の質問文を生成する。

1) Google Suggest/Autocomplete API は Google 検索のサジェストを取得できる一般アクセス可能な API であり、これを用いて構築されたデータセットもある [4] が、利用規約が不明瞭であるため、本研究では利用可能でないものとして扱う。

提案手法は、大規模な生コーパスにも自然な質問を表す文が相当程度含まれており、事前学習を通してモデルがよくある疑問を記憶していることを仮定している。また、クエリフォーカスと情報要求観点を表すキーワード列から質問文を生成するため、質問内容を制御しやすい利点があると考えられる。

本稿では、一般および金融ドメインのキーワードをクエリフォーカスとして25個ずつ指定し、Google検索のサジェストを正解としたT5による情報要求観点の予測性能を定量的に評価する。また、クエリフォーカスと情報要求観点からの質問文生成について、生成結果を定性的に分析する。

2 提案手法

本節では、テキスト生成モデルを用いて自然な質問をマイニングする手法について説明する。提案手法は大まかに以下の3ステップから構成される(図2)。

1. クエリフォーカスを指定し、その情報要求観点を予測させるテンプレート文を生成する。
2. 生成したテンプレート文をT5に入力し、情報要求観点を予測する。
3. LLMを用いて指定したクエリフォーカスと予測された情報要求観点から自然言語の質問文に変換する。

以降、各ステップの詳細について述べる。

2.1 クエリフォーカスの指定

まず、生成する自然な質問の情報欲求の対象を指定し、その情報要求観点を予測させるためのテンプレート文を生成する。具体的には、クエリフォーカスを人手で1つ定め、

- (1) 「Q. {{クエリフォーカス}} の<X>²⁾について教えてください。」

というテンプレートを用いて文を生成する。例えば、クエリフォーカスに「引越す」を指定すると「Q. 引越しの<X>について教えてください。」というテンプレート文が生成される。本稿では単一のテンプレートを用いるが、複数のテンプレートを用いることで予測をより多様化することも可能である。

2) T5はセンチネルトークンと呼ばれるマスクを用いてスパンを予測する事前学習を行っており、センチネルトークンを含むテキストが入力されると、そのマスクされたスパンの予測結果が出力される。<X>はセンチネルトークンを表す。

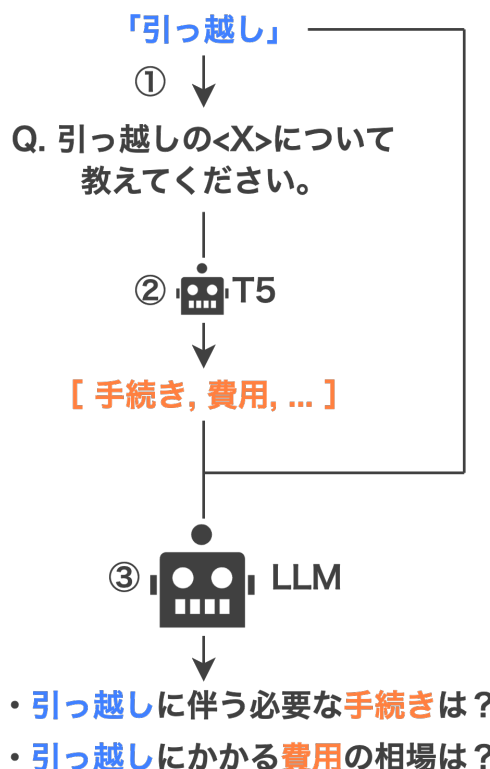


図2 提案手法の概要図。

2.2 情報要求観点の予測

次に、生成したテンプレート文をT5に入力し、指定したクエリフォーカスの情報要求観点を予測する。BERT [10]をはじめとするシングルトークンの補完が可能なマスク言語モデルを用いることもできるが、情報要求観点は「安い時期」といったマルチトークンにもなりうるため、本研究ではマルチトークンの補完が可能なT5を採用する。デコーディングの際は、多様な k 個の情報要求観点の候補を得るために、以下の手順でトークン列を生成する。

1. 先頭のトークンの生成確率を計算し、生成確率が上位 k 件のトークンを取得する。
2. 取得された各トークンの続きを終端記号が出力されるまで生成する。

これは異なる先頭トークンから情報要求観点の候補を生成する制約を課している。

2.3 質問文への変換

最後に、指定したクエリフォーカスと予測した情報要求観点のキーワードをベースに質問文を生成する。LLMに与えるプロンプトのテンプレートを図3に示す。例えば「引越す」および「手続き」とい

指示:
与えられた入力をもとに、それらに関する質問文を1つ作成してください。

入力:
- {{クエリフォーカス}}
- {{情報要求観点}}

図3 キーワードから質問文に変換する際に LLM に与えるプロンプトのテンプレート。

表1 指定したクエリフォーカスの一例。

ドメイン	クエリフォーカス
一般	引っ越し, 海外旅行, 選挙, 入学, 卒業, 就活, 結婚式, 料理, 掃除, 洗濯, スマホ, ...
金融	デリバティブ, 先物取引, 株価収益率, 国際通貨基金, 決算短信, 有価証券報告書, ...

うキーワードが与えられた場合, 「引っ越しに伴う必要な手続きは?」といった質問文を生成させる。なお, 柔軟に質問文を生成させるため, キーワードを生成文に必ず含めるような条件は課していない。

3 実験

提案手法に従って指定したクエリフォーカスに対する情報要求観点を予測し, Google 検索のサジェストを正解として予測性能を定量評価する。また, LLM を用いてクエリフォーカスと情報要求観点から質問文を生成・分析する。

3.1 情報要求観点の予測性能の定量評価

3.1.1 実験設定

クエリフォーカスの指定 ドメインによる予測性能の差を検証するために, 一般および金融ドメインのクエリフォーカスを25個ずつ用意した。表1に本実験で指定したクエリフォーカスの一例を示す。一般ドメインのクエリフォーカスは, 多くの人間にとって情報欲求の対象となりやすい日常生活やライフイベントに関わるキーワードを選定した。金融ドメインのクエリフォーカスは, 経済用語の解説サイト「日経ナレッジバンク」³⁾を参考に選定した。

評価データ 各クエリフォーカスに対する Google 検索のサジェストは Google Suggest/Autocomplete API⁴⁾から取得した。取得の際は末尾に半角スペース

3) <https://www.nikkei4946.com/knowledgebank/>
4) <http://www.google.com/complete/search?gl=jp&hl=ja>

表2 情報要求観点の予測性能。表中の数値は Recall@k を百分率で示している。LLM を用いた情報要求観点の予測性能も予備調査しており, 付録Aを参照されたい。

モデル	一般			金融		
	@1	@10	@100	@1	@10	@100
T5 _{base}	2.5	7.5	19.2	0.0	2.9	12.8
T5 _{large}	2.4	8.8	17.4	1.3	4.6	14.9
T5 _{xl}	2.1	9.4	24.1	0.9	4.1	14.0

スを追加したクエリフォーカスを API に渡し, 得られたサジェストの中で複数のキーワードから成るのは最初のキーワードを抽出した。この結果, 一般ドメインのクエリフォーカスに対する平均サジェスト数は9.56個, 金融ドメインは9.68個であった。

モデル 情報要求観点の予測に用いる T5 は, 日本語 Wikipedia および多言語ウェブコーパス mC4 [11] の日本語サブセットで事前学習されたモデルを用いた。モデルサイズによる影響を検証するため, base⁵⁾・large⁶⁾・xl⁷⁾の3モデルの性能を評価した。

評価指標 サジェストがどの程度予測結果に含まれているかを見るために, 評価指標は Recall@k を用いた。Recall@k は以下の式で計算される。

$$\text{Recall@k} = \frac{1}{N} \sum_{i=1}^N \frac{\text{上位 } k \text{ 個の予測に含まれる正解数}}{\text{正解のサジェスト数}}$$

N は評価データの事例数 (=25) である。k の値は平均サジェスト数を考慮して {1, 10, 100} の3設定で評価した。

3.1.2 実験結果

情報要求観点の予測性能の定量評価結果を表2に示す。一般ドメインのクエリフォーカスに対する予測性能は, モデルサイズが大きいほど高くなる傾向が見られた。また, 下位の予測に正解のサジェストが潜んでおり, 予測をフィルタリング・集約することで性能改善が期待できる。一方で, 金融ドメインは一般ドメインと比較して予測が難しく, 予測性能もモデルサイズと比例する傾向は見られなかった。ウェブコーパスに含まれる金融ドメインの質問文に限りがあるためだと考えられる。

具体的な情報要求観点の予測と正解を表3に示す。予測と正解を比較すると, 部分的に正解である例や意味的に同じだが表現が異なるために不正解である例が散見された。また, 「引っ越し」に対する

5) <https://huggingface.co/retrieva-jp/t5-base-long>
6) <https://huggingface.co/retrieva-jp/t5-large-long>
7) <https://huggingface.co/retrieva-jp/t5-xl>

表4 クエリフォーカスと情報要求観点からの質問文の生成例。

ドメイン	クエリフォーカス	情報要求観点	生成された質問文
一般	引越	手続き	引越の際に必要な手続きは何ですか？
金融	デリバティブ	取引	デリバティブの取引において、最も一般的な形態は何ですか？

表3 T5_{xl} による情報要求観点の予測例。

クエリフォーカス	予測	正解
引越	時期 挨拶 日 手続き 費用 料金 手順 方法 タイミング コソ …	見積もり やること やることリスト 安い時期 挨拶 単身パック 初期費用 手続き 役所 不用品処分
デリバティブ	価格 取引 規制 リスク 取り扱い 売買 分類 仕組み 説明 …	意味 英語 会計処理 時価評価 わかりやすく 現物 種類 本 オプション

「業者」や「荷造り」など、正解にはない妥当な情報要求観点が予測に含まれることもあり、評価指標の改善や人手評価は今後の課題である。他にも、予測結果に「費用」と「料金」といった同義語が多数含まれており、予測の集約が必要だと考えられる。

3.2 質問文の生成・分析

表1の各クエリフォーカスから情報要求観点を10件ずつ予測し、これらのキーワードから2.3節のプロンプトを用いて質問文を生成・分析した。質問文生成に用いるLLMは指示学習済みの13Bパラメータモデル (llm-jp-3-13b-instruct)⁸⁾ [12]を用いた。

表4および付録Bに生成例をいくつか示す。無作為に抽出した50件の生成例を著者が人手評価した結果、44件はキーワード列から想定される妥当な質問文であった。一方で、単なる平叙文を生成するといった指示違反やそもそも予測された情報要求観点が妥当でないために不自然な質問文も散見された。キーワードの意図を予測させてから質問文を生成する多段階のプロンプティングや、より大規模なLLMの利用が改善案として挙げられる。

8) <https://huggingface.co/llm-jp/llm-jp-3-13b-instruct>

4 関連研究

質問生成タスク 質問生成タスクでは文章レベルのテキストや画像を入力し、その内容に対する質問文を生成する。また、解答を同時に入力し、これが正解となるように質問文を生成する設定もある [13]。質問生成タスクの主な対象は答えが短文となる質問文の生成であり [8]、深層学習モデルの訓練・評価データの自動生成や教育応用などを目的とするものが多い。提案手法は、自然な質問の収集を目的とした、キーワードレベルのテキストを入力とする質問文生成手法 [14, 15] と位置付けられる。また、既存研究と比較して、入力キーワードをテキスト生成モデルから収集する点が特徴である。

質問の分類 より多様で汎用的な質問生成システムの実現に向けて、質問のパターンを分類定義する取り組みがいくつか存在する [16, 17]。例えば、Bolotova らは非ファクトイド型の質問を INSTRUCTION・REASON・EVIDENCE-BASED・COMPARISON・EXPERIENCE・DEBATE の6カテゴリに分類定義した。提案手法は単一のテンプレートを用いて情報要求観点を予測するが、これらの質問の分類を参考にテンプレートを多様化することで、広く質問文の収集が可能になると考えられる。

5 おわりに

本研究では、検索クエリログを用いない自然な質問のマイニング手法を提案した。提案手法は、テキスト生成モデルを用いてユーザが指定したクエリフォーカスの情報要求観点を予測し、LLMを用いて指定したクエリフォーカスと予測した情報要求観点から自然言語の質問文を生成するというものである。T5の情報要求観点の予測性能を定量的に評価し、テキスト生成モデルから自然な質問がマイニングできる可能性を示唆した。

今後の課題として、評価の大規模化や情報要求観点の予測の多様化・フィルタリング手法の検討などが挙げられる。また、提案手法に従って自然な質問を広く収集し、これをもとに指示学習（質問応答）データセットを構築する枠組みも検討する。

参考文献

- [1] Tom Kwiatkowski, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur Parikh, Chris Alberti, Danielle Epstein, Illia Polosukhin, Jacob Devlin, Kenton Lee, Kristina Toutanova, Llion Jones, Matthew Kelcey, Ming-Wei Chang, Andrew M. Dai, Jakob Uszkoreit, Quoc Le, and Slav Petrov. Natural questions: A benchmark for question answering research. **Transactions of the Association for Computational Linguistics**, Vol. 7, pp. 452–466, 2019.
- [2] Takuya Uematsu, Hao Wang, Daisuke Kawahara, and Tomohide Shibata. A benchmark suite of Japanese natural questions. In **Proceedings of the 13th Joint Conference on Lexical and Computational Semantics (*SEM 2024)**, 2024.
- [3] 宇津呂武仁. 検索窓から俯瞰する世界. 電子情報通信学会誌, Vol. 99, No. 9, pp. 920–927, Sep 2016.
- [4] Jonathan Berant, Andrew Chou, Roy Frostig, and Percy Liang. Semantic parsing on Freebase from question-answer pairs. In **Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing**, 2013.
- [5] 片山太一, 大塚淳史, 光田航, 齋藤邦子, 富田準二. 相手の発話を深掘りするための質問生成技術. 人工知能学会全国大会論文集, Vol. JSAI2018, pp. 4G103–4G103, 2018.
- [6] Chao-Yi Lu and Sin-En Lu. A survey of approaches to automatic question generation: from 2019 to early 2021. In **Proceedings of the 33rd Conference on Computational Linguistics and Speech Processing (ROCLING 2021)**, 2021.
- [7] Shasha Guo, Lizi Liao, Cuiping Li, and Tat-Seng Chua. A survey on neural question generation: Methods, applications, and prospects. In Kate Larson, editor, **Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24**, pp. 8038–8047. International Joint Conferences on Artificial Intelligence Organization, 8 2024. Survey Track.
- [8] Said Al Faraby, Adiwijaya Adiwijaya, and Ade Romadhony. Review on neural question generation for education purposes. **International Journal of Artificial Intelligence in Education**, Vol. 34, No. 3, pp. 1008–1045, Sep 2024.
- [9] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. **Journal of Machine Learning Research**, Vol. 21, No. 140, pp. 1–67, 2020.
- [10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, 2019.
- [11] Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. mT5: A massively multilingual pre-trained text-to-text transformer. In **Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, 2021.
- [12] LLM-jp. LLM-jp: A cross-organizational project for the research and development of fully open japanese LLMs. **arXiv [cs.CL]**, July 2024.
- [13] Luu Anh Tuan, Darsh Shah, and Regina Barzilay. Capturing greater context for question generation. **Proceedings of the AAAI Conference on Artificial Intelligence**, Vol. 34, No. 05, pp. 9065–9072, Apr. 2020.
- [14] Zhicheng Zheng, Xiance Si, Edward Chang, and Xiaoyan Zhu. K2Q: Generating natural language questions from keywords with user refinements. In **Proceedings of 5th International Joint Conference on Natural Language Processing**, 2011.
- [15] Adarsh Kumar, Sandipan Dandapat, and Sushil Chordia. Translating web search queries into natural language questions. In **Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)**, 2018.
- [16] Rodney Nielsen, Jason Buckingham, Gary Knoll, Ben Marsh, and Leysia Palen. A taxonomy of questions for question generation. 01 2008.
- [17] Valeriia Bolotova, Vladislav Blinov, Falk Scholer, W. Bruce Croft, and Mark Sanderson. A non-factoid question-answering taxonomy. In **Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '22**, p. 1196–1207, New York, NY, USA, 2022. Association for Computing Machinery.

表5 クエリフォーカスと情報要求観点から生成された質問文の追加の例.

ドメイン	クエリフォーカス	情報要求観点	生成された質問文
一般	選挙	日	次の衆議院議員選挙はいつ行われる予定ですか？
	結婚式	ご祝儀	結婚式のご祝儀の相場はいくらですか？
	洗濯	乾燥	洗濯と乾燥のどちらが好きですか？
	介護	看護	介護と看護の違いは何ですか？
金融	先物取引	届出	先物取引の届出はどのように行えばよいですか？
	株価収益率	計算方法	株価収益率（PER）の計算方法について教えてください。
	有価証券報告書	記載事項	有価証券報告書にはどのようなことが記載されているのですか？
	エンゲル係数	変動	エンゲル係数が上昇する場合、一般的に何が原因と考えられますか？

検索エンジンの入力の内、質問の観点を表す2語目以降のことをサジェストと呼びます。与えられた入力に対するサジェストを簡条書きで k 個列挙してください。なお、回答にはサジェスト部分の簡条書きのみを含めてください。
入力: {{クエリフォーカス}}

問文も見られた。

図4 GPT-4o に対してサジェストを予測させるプロンプトテンプレート。

付録

A GPT-4o を用いた情報要求観点の予測

予備調査として、一般ドメインのクエリフォーカス(表1)に対する GPT-4o (2024-05-13)⁹⁾ の情報要求観点の予測性能を定量評価した。使用したプロンプトテンプレートを図4に示す。実験の結果、一般ドメインのクエリフォーカスに対する GPT-4o の情報要求観点の Recall@1 は 3.4, Recall@10 は 18.2, Recall@100 は 21.0 であった。予測の Precision が高いものの、予測数を増やすことによる Recall の増加は限定的であると考えられる。また、T5_{xl} の Recall@100 は 24.1 であることから(表2)、本研究ではよりサジェストを収集できる可能性のある T5 に焦点を当てる。

B 質問文の生成例

クエリフォーカスと情報要求観点から生成された質問文の追加の例を表5に示す。「日」という情報要求観点から「いつ」に関する質問文を生成するといった柔軟な応答が見られた一方で、「洗濯と乾燥のどちらが好きですか?」といった妥当性の低い質

9) <https://openai.com/index/chatgpt/>