

クリックログと小規模高品質データを併用した E コマースクエリ意図分類モデルの精度向上

田爪聡¹ 伊奈拓郎¹ 馬縹美穂¹ 石原敬大¹ 鍛冶伸裕¹

¹LINE ヤフー株式会社

{stazume,tina,mimatsun,takishih,nkaji}@lycorp.co.jp

概要

本研究では、E コマースにおけるクエリ意図分類の精度向上を目的とする。提案手法では、クリックログを利用して大規模なデータセットを構築し、それを人手ラベルデータで補完する方法を採用した。さらに、これらのデータを統合的に活用することで、モデルを段階的に学習させる新しいアプローチを提案する。実験の結果、クリックログによる大規模データセットを信頼性の高い小規模な人手データで補完することで、従来手法を上回る分類性能を示した。

1 はじめに

近年、E コマースの市場規模は拡大しており [1]、膨大な商品の中から必要な商品を検索する技術が、ますます重要になっている。多くの E コマースサイトでは、ユーザが探したい商品に関する情報をクエリとして入力し、検索する手段が提供されている。その一方で、多くの検索手段は表層的な文字列マッチングに依存しており、ユーザの検索意図を十分に捉えることができない場合が多い。例えば、ユーザが「メロンソーダ」と検索した場合、飲料カテゴリが適切であるにも関わらず、E コマースサイトには多種多様な商品が存在するため、関連性の低い商品、例えばメロンソーダ味の歯磨き粉が検索結果に表示されることがある。クエリ意図を正確に理解し、適切な商品や情報を提示することは、ユーザ体験の向上と収益拡大に直結する。このような問題を解決するための手段として、クエリがどのカテゴリに属するかを予測するクエリ意図分類が研究されている [2, 3, 4, 5, 6, 7]。本研究では、クエリがどのカテゴリに属するかを予測するクエリ意図分類のタスクに取り組む。なお、「ワンピース」というクエリのように、ファッションカテゴリやコミックカテゴ

リなど複数のカテゴリに属するケースも存在するため、本研究のクエリ意図分類はマルチラベル分類のタスクとして扱う。

クエリ意図分類を実現するためには、機械学習モデルの活用が一般的である。その際、クエリとそれに対応する意図ラベルが付与された大規模データセットは、モデルの高い汎化性能を達成する上で重要である。しかし、このような大規模かつ高品質なデータセットを構築することは容易ではない。たとえば、ヤフーショッピングにおいては、クエリ意図候補のカテゴリは 1 万種類以上存在し、クエリごとに正確なラベル付けを行うには膨大な作業コストが必要となる。さらに、1 つのクエリに複数の意図が関連付けられる可能性があるため、すべてのクエリについて全カテゴリと網羅的に対応付けることは現実的に困難である。

多くの既存研究では、この課題を解決するために、クリックログを活用して大規模なデータセットを自動構築するアプローチが採用されている [2, 3, 4, 6, 7]。クリックログは、ユーザの検索クエリとクリック行動を記録したデータであり、ユーザの意図を反映した有益な情報源である。一方で、誤クリックや不正確なカテゴリ情報が含まれる場合が多く、データにノイズや不整合が生じることが課題となる。その結果、クリックログに基づくデータセットのみを利用すると、ノイズの影響でモデルの性能が低下するリスクがある。

そこで本研究では、クリックログ由来の大規模低品質データに加えて、人手でラベル付けした小規模高品質データを利用することを提案する。後者のデータはクラウドソーシングを利用して作成した。具体的には、評価者に検索クエリ、商品名、商品のカテゴリ情報、および商品の詳細ページを提示し、商品のカテゴリがクエリ意図と一致しているかどうかを評価させた。これにより、少量ながらも信頼性

の高いクリーンなデータを作成することに成功した。モデル学習時には、クリックログを用いてファインチューニングを行い、その後人手ラベルデータセットを用いて追加でファインチューニングを行うことで、モデルの分類精度を向上させることを目指した。このアプローチの有効性を実験的に検証した結果、少量のクリーンデータを活用することで、クリックログのみを用いた場合に比べて精度が向上することを確認した。

2 関連研究

クエリ意図分類は、E コマースサイトにおける検索精度の向上において重要な課題とされ、多くの研究が行われてきた。その中でも、学習データの構築に伴うアノテーションコストの高さが大きな障壁となっており、この課題を克服するために、クリックログを利用して学習データを生成する手法が一般的に採用されている [2, 3, 4]。しかし、クリックログにはノイズが含まれるため、本研究ではこれに加えて、人手でラベル付けしたデータも組み合わせる利用し、意図分類の精度向上を目指す。

一方、クリックログに加えて商品データを活用する転移学習のアプローチも注目を集めている。たとえば、Skinner らは、大量の「商品名-カテゴリ」データを用いて LSTM モデルを事前学習し、その後、少量の「クエリ-カテゴリ」データを用いた転移学習を行うことで、分類精度を向上させることに成功している [5]。この手法は、商品データの豊富さを活かしつつ、クエリデータの不足を補完する効果がある。

さらに近年では、BERT[8] のような事前学習モデルをクエリ意図分類に応用する研究も進んでいる。G Di Fabbrizio らは、DistilBERT[9] と sparsemax 損失 [10] を組み合わせることで、多クラス分類や疎なラベル分布に対応可能なモデルを構築した [6]。また、Yiming Qiu らは、BERT モデルの事前学習に一般的な文章ではなく、ユーザ検索クエリと商品名部分文字列を利用するアプローチを提案し、E コマース特有のテキスト特性を活かして分類性能を大幅に向上させた [7]。

本研究で提案するアプローチは、こうした転移学習や事前学習と併用可能なものになっている。本論文では事前学習済み BERT モデルを用いるが、そのほかの事前学習済みモデルを使うことも可能である。

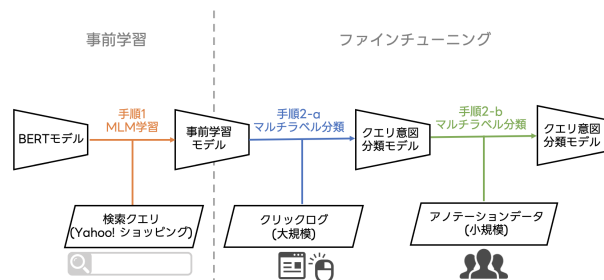


図1 モデル学習の概要図

3 提案手法

3.1 モデル学習の流れ

本研究では、クエリ意図分類の基盤モデルとして、Transformer アーキテクチャ [11] に基づく事前学習モデルである BERT モデル [8] を採用した。モデル学習の全体的な流れを図1に示す。

事前学習には、検索クエリを用いてマスク化言語モデルを学習した。このアプローチにより、汎用的な BERT モデルと比較して、検索クエリ特有のデータに適応した表現を効率的に学習できる。

次に、事前学習を終えた BERT モデルに対してファインチューニングを行った。本研究では、データの特性に応じてファインチューニングデータセットを段階的に分けるアプローチを採用した。まず、クリックログを用いた大規模データで粗学習を実施し、その後、小規模ながら高品質な人手ラベルデータを用いて精密学習を行う。この手法により、クリックログの大規模性を活かしつつ、ノイズの影響を人手データで軽減した学習が可能となる。

3.2 クリックログデータによる学習

第一段階として、クリックログから自動生成した大規模低品質な学習データ (4.1 節) を用いた BERT モデルの学習を行う。このステップでは、BERT モデルが多くのクエリとすべてのカテゴリ間の関係を学習することを目的としている。具体的には、BERT モデルの出力層において、[CLS] トークンの隠れ状態に対して一層の線形変換を適用し、その出力にシグモイド関数を用いることで各カテゴリに対する二値分類モデルを学習する [12]。すなわち、カテゴリ c に対してクエリ q が適合するか否かを確率値として出力する形式で学習を行った。

3.3 人手ラベルデータによる学習

第二段階では、第一段階で学習済みのモデルの重みを初期値として、人手ラベルデータを用いた追加のファインチューニングを行った。このデータセットは、クエリ q とカテゴリ c の適合性が人手でアノテーションされており、ノイズが少なく高品質である (4.2 節参照)。

学習時には、クエリ q とカテゴリ c のペアを BERT モデルに入力し、[CLS] トークンの出力を基に線形層を介して損失を計算した。この損失は、対応するカテゴリ c のラベル (適合または不適合) に基づいて定義され、2 値分類タスクとしてモデルを最適化した。この段階で、モデルのパラメータを最適化し、より精密な分類性能を実現した。

これらの二段階ファインチューニングにより、クリックログの大規模性と人手ラベルデータの高品質性を組み合わせて活用することが可能となった。その結果、ノイズに対する耐性を持つ高精度なクエリ意図分類モデルの構築を実現した。

4 データの構築手順

4.1 クリックログベースのデータの作成

大量のクエリに対して、それぞれの意図するカテゴリを正確にラベル付けする作業は非常に困難である。特にヤフーショッピングでは、取り扱うカテゴリ数が 1 万を超えており、機械学習モデルの学習に必要なデータ量は膨大になる。このような状況下では、一つのクエリが複数のカテゴリに属する可能性があり、これを人手でラベル付けすることは極めて高いコストと労力を伴う。

本研究では、Chang[2] らのアプローチを参考に、クリックログを用いて正解データを自動的に作成する。クリックログを用いることで、クエリとカテゴリの対応関係を効率的に取得し、それを機械学習モデルの学習データとして利用する。この方法により、人手によるラベル付けを必要としない、効率的かつ大規模なデータセットの構築が可能となる。

ヤフーショッピングでは、ユーザが検索したクエリと、その際にクリックされた商品の情報がログとして記録される仕組みが構築されている。各商品には事前にカテゴリが付与されており、これにより 1 回のクリックからクエリとカテゴリのペア (q, c) を獲得できる。このようなペアを大量に収集し、クエ



図2 アノテーションデータの作成フロー

リ q 毎にカテゴリ c のクリック頻度を集計することで、カテゴリ分布を計算することが可能である。一定回数以上のクリックがあったクエリカテゴリペアを正例とすることによって、各クエリのカテゴリ意図をモデルが効果的に学習できるデータセット D_{log} を構築する。

4.2 クラウドソーシングを利用したデータの作成

前節では、クリックログを用いたデータセット構築手法を紹介した。しかし、クリックログのみに依拠したデータセットでは、ユーザの誤クリックや、商品に紐付けられたカテゴリ情報の誤付与など、ノイズの混入を完全に防ぐことは難しい。

そこで本研究では、クリックログによるデータセットを補完するため、人手によるアノテーションを活用した高品質なデータセットを作成した。クエリに対して適切なカテゴリを 1 万種類以上の候補から直接選択させるのは現実的ではないが、特定の商品が提示された場合にそのクエリ意図と商品情報が適合しているかを判定するタスクは比較的容易である。このことを利用し、クエリとヒットする商品のペアを評価者に提示することで、間接的にクエリと商品カテゴリの対応を得るデータを構築した。

データセットの作成にはクラウドソーシングを用い、評価者には検索クエリ、商品名、カテゴリ情報、商品の詳細ページを提示した。評価者はこれらの情報に基づき、クエリと与えられた商品のカテゴリがクエリに対して妥当かどうかをラベル付けした。商品名や詳細ページの参照が可能な設定により、ラベル付けの精度を高めている。

これらのアノテーション結果を基に、クエリと商品カテゴリの適合性を基準にラベル付けされたデー

表 1 評価結果

学習データ	おすすめ順		安い順	
	全クエリ	テイルクエリ	全クエリ	テイルクエリ
D_{log}	0.962	0.895	0.966	0.989
$D_{log} + D_{annot}$	0.965	0.909	0.970	0.990

タセットを構築した。本研究では、この高品質な人手ラベルによるデータセット D_{annot} を活用し、クリックログデータと統合することで分類性能の向上を検証する。

5 実験と考察

5.1 実験設定

本研究では、次の 2 種類のモデルを比較することで、二段階ファインチューニングが分類精度に与える影響を評価した。一つ目は、 D_{log} のみを用いた一段階目のファインチューニングのみで学習を完了させた「クリックログモデル」である。二つ目は、クリックログモデルに対して、人手ラベルデータセット D_{annot} を用いた二段階目のファインチューニングを適用した「ハイブリッドモデル」である。

モデルの詳細 モデルには、Transformer アーキテクチャを基盤とする二層構造の BERT モデルを採用した。トークナイザとして SentencePiece[13] を使用し、Unigram モデル [14] で入力テキストをトークン化した。入力シーケンスの最大長は 32 トークンに設定し、短い検索クエリに適した設計とした。

分類タスクのラベルにはヤフーショッピングのカテゴリを使用した。本タスクは、13,000 以上のクラスを持つマルチラベル分類問題として定義している。

学習データに関する補足 学習データの収集では、ヤフーショッピングの並び替え機能のうち、売上などを考慮した「おすすめ順」と価格順で並び替えられた「安い順」のデータを利用した。テイルクエリ（検索頻度の低いクエリ）はクリックログが少ないため、ノイズの影響を受けやすい。そのため、クエリのサンプリングではテイルクエリの割合を増やすよう調整を行い、データのバランスを図った。人手ラベルデータセット D_{annot} は 26,000 レコードを含み、これを学習データと評価データに 9:1 の割合で分割して使用した。

5.2 実験結果と考察

本研究における実験結果を表 1 に示す。評価は、評価データ全体およびテイルクエリのみに行った評価を実施した。性能指標として、precision-recall 曲線の AUC (Area Under the Curve) を採用した。

表 1 に示された結果から、いずれのソート順においても、 D_{log} と D_{annot} を用いて学習したハイブリッドモデルが D_{log} のみで学習したクリックログモデルと比較して高い性能を示した。またテイルクエリのみを用いた評価でもその優位性は変わらなかった。

さらに、個別事例の比較により、モデル間の特性を詳細に分析した。推論結果の詳細については、付録 A を参照されたい。例えば、クエリ「ボールペン画」に対する結果を分析したところ、カテゴリ「本、雑誌、コミック > 趣味 > イラスト、カット > イラスト」が「適合」としてアノテーションされている。このケースにおいて、クリックログモデルは「キッチン、日用品、文具 > 文具、ステーションナリー > 筆記用具 > ボールペン」といった文房具カテゴリを返却しているのに対し、ハイブリッドモデルは正解ラベルを正確に返却することに成功している。

また、クエリ「醤油 野沢菜の漬け方」に対する結果を分析したところ、カテゴリ「食品 > 漬物、佃煮、ふりかけ > 漬物」が「不適合」とラベル付けされているにもかかわらず、クリックログモデルはこの誤ったカテゴリを返却した。一方で、ハイブリッドモデルは「本、雑誌、コミック > 生活 > 家庭料理 > 家庭料理」といった料理本カテゴリを主に返却しており、クエリの表層的な表現に依存せず、正解に近いクエリ意図を的確に捉えていることが示唆された。

6 まとめ

本研究では、大規模なクリックログと小規模な人手ラベルデータを組み合わせた二段階の学習アプローチを提案した。実験により、提案手法は各条件で従来手法の精度を上回ることを示した。さらに、個別の事例分析を通じて、提案手法が既存手法と比較してクエリ意図をよりの確に捉えた分類結果を返却することが確認された。本研究で提案したフレームワークは、E コマースに限らず、さまざまなクエリ意図解釈モデルの性能向上に幅広く適用可能である。

参考文献

- [1] 経済産業省 商務情報政策局情報経済課. 電子商取引に関する市場調査 報告書, 2023.
- [2] Yiu-Chang Lin, Ankur Datta, and Giuseppe Di Fabbri. E-commerce product query classification using implicit user's feedback from clicks. In **2018 IEEE International Conference on Big Data (Big Data)**, pp. 1955–1959, 2018.
- [3] Xianjing Liu, Hejia Zhang, Mingkuan Liu, and Alan Lu. System design of extreme multi-label query classification using a hybrid model. In **eCOM@SIGIR**, 2019.
- [4] Lvxing Zhu, Hao Chen, Chao Wei, and Weiru Zhang. Enhanced representation with contrastive loss for long-tail query classification in e-commerce. In Shervin Malmasi, Oleg Rokhlenko, Nicola Ueffing, Ido Guy, Eugene Agichtein, and Surya Kallumadi, editors, **Proceedings of the Fifth Workshop on e-Commerce and NLP (EC-NLP 5)**, pp. 141–150, Dublin, Ireland, May 2022. Association for Computational Linguistics.
- [5] Michael Skinner and Surya Kallumadi. E-commerce query classification using product taxonomy mapping: A transfer learning approach. In **eCOM@SIGIR**, 2019.
- [6] Giuseppe Di Fabbri, Evgeny Stepanov, and Filippo Tessaro. Extreme multi-label query classification for e-commerce. In **The SIGIR 2024 Workshop on eCommerce, Washington, DC, USA**, 2024.
- [7] Yiming Qiu, Chenyu Zhao, Han Zhang, Jingwei Zhuo, Tianhao Li, Xiaowei Zhang, Songlin Wang, Sulong Xu, Bo Long, and Wen-Yun Yang. Pre-training tasks for user intent detection and embedding retrieval in e-commerce search. In **Proceedings of the 31st ACM International Conference on Information & Knowledge Management**, pp. 4424–4428, 2022.
- [8] Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of naacl-HLT**, Vol. 1, p. 2. Minneapolis, Minnesota, 2019.
- [9] V Sanh. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. **arXiv preprint arXiv:1910.01108**, 2019.
- [10] Andre Martins and Ramon Astudillo. From softmax to sparsemax: A sparse model of attention and multi-label classification. In **International conference on machine learning**, pp. 1614–1623. PMLR, 2016.
- [11] A Vaswani. Attention is all you need. **Advances in Neural Information Processing Systems**, 2017.
- [12] Hamada M Zahera, Ibrahim A Elgendy, Richa Jalota, Mohamed Ahmed Sherif, and E Voorhees. Fine-tuned bert model for multi-label tweets classification. In **TREC**, pp. 1–7, 2019.
- [13] Taku Kudo and John Richardson. SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In Eduardo Blanco and Wei Lu, editors, **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations**, pp. 66–71, Brussels, Belgium, November 2018. Association for Computational Linguistics.
- [14] Taku Kudo. Subword regularization: Improving neural network translation models with multiple subword candidates. In Iryna Gurevych and Yusuke Miyao, editors, **Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 66–75, Melbourne, Australia, July 2018. Association for Computational Linguistics.

A 具体的な出力事例

事例 1

クエリ：「ボールペン画」

カテゴリ：本、雑誌、コミック＞趣味＞イラスト、カット＞イラスト

正解ラベル：適合

モデル	出力カテゴリ	予測ラベル
クリックモデル	キッチン、日用品、文具＞文具、ステーションナリー＞筆記用具＞ボールペン＞ボールペン キッチン、日用品、文具＞文具、ステーションナリー＞筆記用具＞万年筆＞万年筆 楽器、手芸、コレクション＞美術、工芸品＞絵画＞日本画 楽器、手芸、コレクション＞画材、アート用品＞その他画材、アート用品 キッチン、日用品、文具＞文具、ステーションナリー＞筆記用具＞マーカー、蛍光ペン＞マーカー	不適合
ハイブリッドモデル	本、雑誌、コミック＞芸術＞絵画技法書＞絵画技法 本、雑誌、コミック＞芸術＞絵画、作品集＞絵画、作品集全般 楽器、手芸、コレクション＞美術、工芸品＞絵画＞日本画 本、雑誌、コミック＞芸術＞絵画、作品集＞絵画、作品集その他 本、雑誌、コミック＞趣味＞イラスト、カット＞イラスト	適合

表 2 適合ペアに対する出力結果

事例 2

クエリ：「醤油 野沢菜の漬け方」

カテゴリ：食品＞漬物、佃煮、ふりかけ＞漬物

正解ラベル：不適合

モデル	出力カテゴリ	予測ラベル
クリックモデル	食品＞漬物、佃煮、ふりかけ＞漬物 食品＞漬物、佃煮、ふりかけ＞その他漬物、佃煮、ふりかけ 食品＞漬物、佃煮、ふりかけ＞佃煮 食品＞調味料、料理の素、油＞その他調味料、料理の素、油 食品＞惣菜、料理＞その他惣菜、料理	適合
ハイブリッドモデル	本、雑誌、コミック＞生活＞家庭料理＞家庭料理 本、雑誌、コミック＞生活＞家庭料理＞家庭料理全般 本、雑誌、コミック＞生活＞家事、マナー＞くらしの知恵、節約 本、雑誌、コミック＞生活＞料理その他＞料理その他全般 本、雑誌、コミック＞生活＞家庭料理＞和食	不適合

表 3 不適合ペアに対する出力結果