

擬似順序付け合成データによる 具体性を考慮した密検索モデルの構築

P09

福井智矢、山本岳洋、湯本高行（兵庫県立大学）

レビュアーA

吸引力はそんなにありませんが、ちょこちょこお掃除するのに便利です

レビュアーB

カーペットに絡んだ犬の毛を吸うのは苦戦しましたが、手軽に掃除できて便利です

閲覧者

吸引力はどれくらい？
もっと具体的に書いたレビューはない？

それくらいの吸引力なら良さそう

目的：抽象的なレビューからより具体的なレビューを検索

問題意識

Q. 今の検索技術では何が問題？

A. 具体性を考慮して検索できない

疎検索(例: BM25)：クエリと文書との単語の一致度で検索

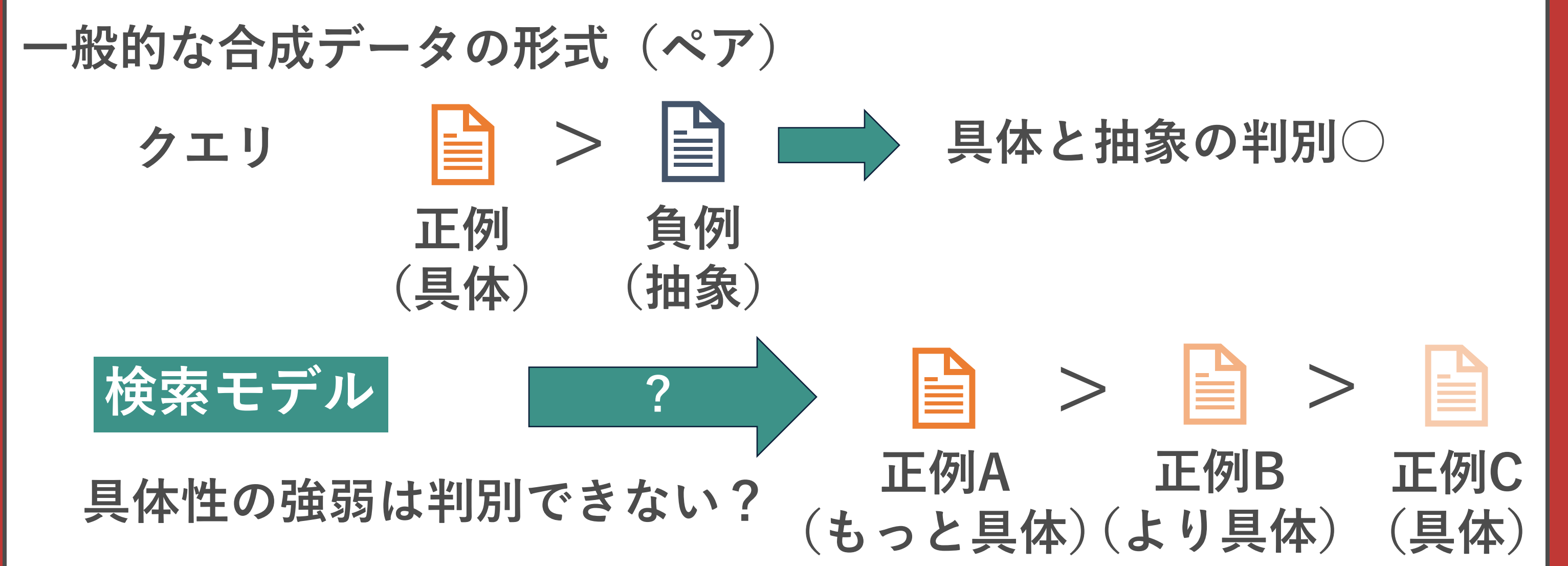
→ クエリに具体的な単語が必要

密検索(例: DPR)：クエリと文書のベクトル表現間の類似度で検索

具体性を考慮したベクトル表現の学習が行われていない
具体性を学習できる学習データセットがない

Q. 具体性を考慮した合成データ作成では何が問題？

A. 具体性の強弱（相対関係）を反映していない



解決策：具体性を考慮した合成データを作成

解決策：具体性による擬似順序付けを行う

提案

問題定義

抽象的なレビューと知りたいことから
より具体的なレビューを検索

レビューにはいろいろな観点があり、検索者がどの観点についてより具体的に知りたいと考えているのかわからない

→ 抽象的なレビューに含まれる知りたいことを
入力として受け取る

知りたいこと → 抽象的なレビュー → 検索モデル → より具体的なレビューランキング

合成データ生成方法

抽象的なレビュー → → 知りたいこと
擬似順序付けしたレビュー

```
# Task
Given an abstract review about a product, your mission is to:
1. Extract a target phrase from the review.
2. Write four different reviews, each with a different level of concreteness relative to the abstract review.

Specifically, you should write a review for each of the following concreteness levels:
- "concrete review": a review that is concrete relative to the abstract review.
- "highly concrete review": a review that is more concrete than the concrete review.
- "perfectly concrete review": a review that is more concrete than the highly concrete review.

## Input
product: {product}
abstract review: {review}

## Output
```

以下のようなランキングが得られるように学習

- 1: Perfectly
- 2: Highly
- 3: Concrete
- 4: 抽象的なレビュー

↑ 具体度

実験

擬似順序付け合成データを用いた学習により
ランキング上位での検索精度が向上

ruri-v3-reranker-310m	nDCG@1	nDCG@2	nDCG@10
ベースモデル	0.380	0.349	0.367
InfoNCE (ペア)	0.380	0.318	0.336
InfoNCE (擬似ペア)	0.400	0.315	0.281
提案手法 Lambda (順序付け)	0.420	0.350	0.338

実験設定

テストデータセット
楽天レビューデータセット (家電)
50件の (レビュー、知りたいこと) を用意

(レビュー、知りたいこと) の組み合わせに対して
各商品のレビューがより具体的かのアノテーションを実施 → nDCGの利得

合成データ
使用モデル: GPT-4o mini
データサイズ: 10,000 → 訓練データ: 8,000 + 検証データ: 2,000

ペアデータセット: 正例となる具体的なレビューを1件生成
擬似ペアデータセット: 順序付けデータをペア形式に変換

損失関数
ペアワイズ学習: InfoNCE リストワイズ学習: Lambda (nDCGベース)

実際に生成されたレビュー
抽象的なレビュー

…ただ、想像していたよりも吸引力が弱くて残念でした。

擬似順序付け合成レビュー → 具体的な合成レビュー間の差を活用

- 1: 掃除機を使って子供がこぼしたお菓子の屑を掃除したところ、掃除機を最大の吸引力にしても、特にカーペットの奥に入り込んだ細かいゴミや、ペットの毛が全く取れず、何度も繰り返し掃除する羽目になりました。…
- 2: この掃除機を使ってみると、カーペットの隅にある細かいゴミを取り切れず、特にペットの毛が残ってしまうことが多かったです。…
- 3: この掃除機は、カーペットの汚れを取るのに時間がかかり、期待していたほどの吸引力が感じられませんでした。特に、ペットの毛が絡まってしまい、何度も掃除機をかけないとイケませんでした。

Q. なぜランキング下位では性能を向上させることができない？
(仮説) トピックの適合性の判別能力が低下したから？

負例：知りたいことについては書いてあるが抽象的
知りたいことについて書いてあるレビューを検索しにくくしている

今後の展望

- トピックの適合性の判別能力を低下させない学習
- 他の合成データ生成方法の模索