

大規模言語モデルを用いたクイズ生成アルゴリズムの作成

早稲田大学高等学院1年 高橋彰仁

要旨

GPT3.5・4とGoogle Custom Search APIを用いて、ユーザーが入力したトピックの情報を収集し、それらの情報からクイズを生成するアルゴリズム「Barista」を作成した。その結果、Baristaが生成したクイズは、定量的な評価にて**86%の割合で人間に近い質**を示したため、人間によるクイズ作成を代替できるアルゴリズムを作ることができた。

研究背景

クイズアプリでのトピック不足を感じた

↓
多種多様なトピックを扱い、
興味のあるクイズを楽しめるようにしたい

Baristaの設計/フローチャート/動作例

Baristaは、以下の4つのアルゴリズムの組み合わせである。

情報収集アルゴリズム:

プロンプトとその応答を用いて、トピックの詳細かつ大量の情報を検索して収集

クイズ生成アルゴリズム:

収集データをテキストファイルに保存し、GPT-4で分割済み収集データからクイズ生成

短期記憶アルゴリズム:

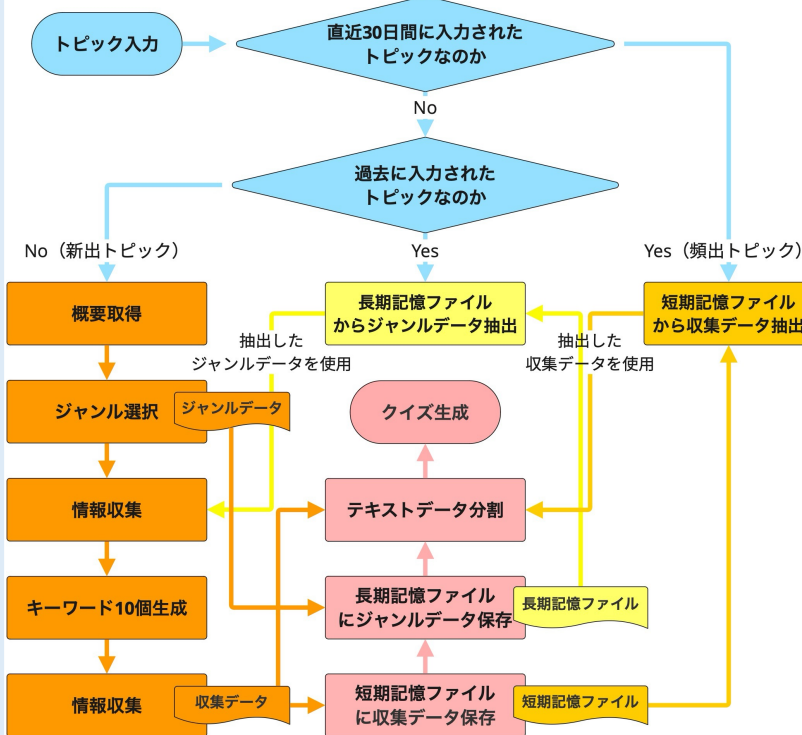
直近30日間に収集したデータをテキストファイルに保存し、入力トピックに応じて使用

長期記憶アルゴリズム:

今までのトピックのジャンルをテキストファイルに保存し、入力トピックに応じて使用

また、この4つのアルゴリズムを入力されたトピックの頻出度合いに応じて使い分けるため

- Baristaを使用すればする程、記憶が増えて費用対効果が高くなる
- 最小限のリソースで直近30日間の最新情報からクイズを生成できるという利点がある。



Baristaが生成したクイズの評価

学校の生徒45人に、記載されたクイズ(4選択肢問題)の・正確性・明瞭性・関連性・多様性・娯楽的価値を評価する・作り手(人間/Barista/GPT-4のいずれか)を予想するアンケートを実施した。

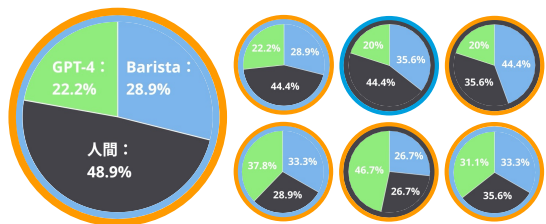
1. 5要素の評価(質を評価)

各作り手のクイズに対して5段階評価をした。5に近い程「高い」、0に近い程「低い」評価である。

	正確性	明瞭性	関連性	多様性	娯楽的価値	総合
人間	5.00	4.47	4.20	3.91	4.00	4.32
Barista	5.00	3.98	4.13	3.53	3.56	4.04
GPT-4	0.50	4.20	4.33	3.40	3.18	3.12

結果、人間が最も高い質のクイズを作成していた。Baristaは詳細な情報を参照するため、**複雑でトピックから飛躍したクイズ(マニアックなクイズ)を生成する傾向**があることから、明瞭性・関連性で人間やGPT-4を下回ってしまったが、正確性・多様性・娯楽的価値でGPT-4を上回ったことで、総合では、GPT-4を大きく上回り、人間に近いスコアを出した。

2. 作り手の予想(質の安定性を評価)



(円グラフ内枠)作り手:人間(黒) Barista(青)
(外枠)評価:人間と同等(オレンジ) 人間未満(青)

予想結果の比率が下記のようにあれば、作り手の見分けがつかない、つまりBaristaが生成したクイズは人間が作成したクイズと同等の質であると言える。

作り手が人間の場合: $人間 \approx 人間$
作り手がBaristaの場合: $人間 > \approx Barista$

結果、7問中6問が上記の比率になった。よって、Baristaは86%の割合で人間に近い質のクイズを生成できると言える。1問が上記の比率にならなかった要因は、回答者が上記のBaristaの傾向に気づいたからだと考えた。

Barista 必ず<ジャンルとジャンルキーワード>の中から、キーワードと情報に当てはまるジャンルとジャンルキーワードを最大2つ選択してください。
キーワード:{topic}
情報:{top_result} ... (後略)

GPT-4 選択したジャンル: 音楽
ジャンルキーワード: 「King Gnu とは」
選択したジャンル: アーティスト
ジャンルキーワード: 「King Gnu おすすめの曲」

GPT-3.5 {topic}についての入力された情報から、重要な{topic}に関する人、もの、事柄を必ず10個挙げてください。
入力された情報:{subkey10_data}

あなたは選択肢付き問題形式で、入力された情報から{topic}に関する面白い正しい問題を10問作成してください。以下の表示形式を厳守してください。
表示形式(厳守):
Q: 問題文
(A) 選択肢A (B) 選択肢B (C) 選択肢C (D) 選択肢D A: 正解の選択肢
入力された情報:{input_text}

Q: King Gnuのドラマー、勢喜遊が使用しているドラムメーカーはどこでしょう?(A) Yamaha Drums (B) Pearl Drums (C) TAMA Drums (D) DW Drums
A: (C) TAMA Drums ... etc (10問生成)



結論・今後の展望

評価結果より、Baristaは、人間によるクイズ作成を代替できるアルゴリズムであると言える。今後は、Barista内のGPT-3.5・4をプロンプト/ファインチューニングして、ユーザーが求める難易度に合わせてクイズ生成の傾向の強さを調整可能にし、さらに、クイズの質向上/処理速度向上/低コスト化を進めたい。また、Baristaを軽量化や著作権回避などさせたBarista PublicをバックエンドとするWebサービスをリリースし、トピックが無制限の新たなクイズ体験を提供したい。

参考文献・付記

- Baristaソースコードと補足説明
(<https://colab.research.google.com/drive/1MZ8q18I-Gsp56ku9yRkL1NvYk1XMOe2?usp=sharing>)
 - Barista Publicソースコードと補足説明
(https://colab.research.google.com/drive/1OrSQvRAoDd1raMUOX4ok69_0wTm_Ka-z?usp=sharing)
- 本ポスターは2023年度早稲田大学高等学院同窓会学術奨励金の研究成果の一部である。

