

# 誤答選択肢の作成方略を与えた大規模言語モデルに基づく多肢選択式問題の誤答選択肢自動生成手法

## Automatic Distractor Generation in Multiple-Choice Questions Using Large Language Models Given Distractor Construction Strategies

永井 優成<sup>\*1</sup>, 宇都 雅輝<sup>\*1</sup>  
Yusei Nagai<sup>\*1</sup>, Masaki Uto<sup>\*1</sup>  
<sup>\*1</sup> 電気通信大学

<sup>\*1</sup>The University of Electro-Communications  
Email: {nagai, uto}@ai.lab.uec.ac.jp

**あらまし：** 多肢選択式問題自動生成における重要タスクの一つとして、誤答選択肢の生成が挙げられる。従来の大規模言語モデルを用いた誤答選択肢生成では、選択肢生成に関わる専門知識を明示的に扱わず、モデルが持つ知識のみに頼った生成を行っている。そこで本研究では、誤答選択肢を作る際の具体的な考え方（誤答選択肢作成方略）をモデルに与えることで、特定の方略に沿った誤答選択肢を大規模言語モデルに生成させる手法を提案する。

**キーワード：** 読解問題, 誤答選択肢生成, 深層学習, 大規模言語モデル, 言語生成

### 1 はじめに

読解力育成方法の一つとして、学習者に様々な文章を読ませ、それぞれの文章に関連する多様な読解問題に取り組みさせる方法が知られている。しかし、様々な文章に関する大量の問題を人手で作成することは、多くの労力を要する。この問題を解決する方法の一つとして、人工知能技術を用いて多肢選択式の読解問題を自動で生成する読解問題自動生成技術が多数提案されている。

多肢選択式問題自動生成における重要タスクの一つとして、誤答選択肢の作成が挙げられる。誤答選択肢の生成品質の向上を目指した先行研究としては、大規模言語モデルを用いて誤答選択肢の候補を生成し、フィルタリングやランキング技術により不適切な候補を除外する方法 [1] や、大規模言語モデルにプロンプトを数段階に分けて与えて、問題と誤答選択肢を段階的に生成する方法 [2] などが提案されている。しかし、これらの手法は、選択肢作成に関わる専門知識を明示的に扱わず、大規模言語モデルが持つブラックボックスな知識のみに頼った生成を行っている。他方で、人間の専門家が誤答選択肢を作る際の具体的な考え方（以降、「誤答選択肢作成方略」と呼ぶ）を明示的に生成モデルに考慮させることができれば、より高品質な誤答選択肢の生成が期待できる。

そこで本研究では、誤答選択肢の生成品質を高めるために、誤答選択肢の具体的な作成方略を整理した上で、大規模言語モデルに特定の作成方略に沿った誤答選択肢を生成させる手法を提案する。

### 2 提案手法

本研究では、誤答選択肢作成方略を考慮して大規模言語モデルに基づく誤答選択肢生成を行う手法を提案する。提案手法は以下の 4 ステップから構成されている。ステップ 1~3 は誤答選択肢生成の準備段階にあたり、

ステップ 4 が実際の生成手順に該当する。

#### 2.1 誤答選択肢作成方略の分類

はじめに、多肢選択式問題に関するテスト理論の先行研究を参考に、誤答選択肢作成方略を整理する。最終的な誤答選択肢作成方略のリストを表 1 に示す。

#### 2.2 方略付き問題データセットの構築

次に、大規模言語モデルの一つである Llama3 を用いて、既存の多肢選択式問題のデータセット  $D = \{(c_n, q_n, a_n, d_n) | n \in \mathcal{N}\}$  に作成方略のラベルを付与する。ここで、 $N$  を問題数として  $\mathcal{N} = \{1, \dots, N\}$ 、 $c_n$  は  $n$  番目の問題に対応する読解文、 $q_n$  は  $n$  番目の問題、 $a_n$  はそれに対応する正答選択肢、 $d_n = \{d_{nk} | k \in \mathcal{K}\}$  ( $\mathcal{K} \in \{1, \dots, K\}$ ) は  $K$  個の誤答選択肢を意味する。まず、データセット  $D$  中のデータ  $(c_n, q_n, a_n, d_n)$  を誤答選択肢ごとに分割して、 $D' = \{(c_n, q_n, a_n, d_{nk}) | n \in \mathcal{N}, k \in \mathcal{K}\}$  とし、 $D'$  のそれぞれのデータに対して Llama3 にどの誤答選択肢作成方略が対応しているかを判定させ、方略ラベル  $y_{nk} \in \{1, 2, 3\}$  を付与したデータセット  $D_{\text{sample}} = \{(c_n, q_n, a_n, d_{nk}, y_{nk}) | n \in \mathcal{N}, k \in \mathcal{K}\}$  を構築する。以上の手続きにより構築された方略付き問題データセット  $D_{\text{sample}}$  を  $D_{\text{train}}$  と  $D_{\text{test}}$  に分割し、以降のステップで用いる。

#### 2.3 誤答選択肢作成方略選定機の構築

このステップでは、任意の読解文と問題、正答選択肢が与えられたときに、どの誤答選択肢作成方略を利用すべきかを判定させるモデルを構築する。本研究ではこのモデルを「誤答選択肢作成方略選定機」と呼び、BERT (Bidirectional Encoder Representations from Transformers) モデルを利用して実現する。

#### 2.4 誤答選択肢の生成

提案手法による誤答選択肢の生成は、次の手順で行う。まず、任意の読解文  $c$ 、問題  $q$ 、正答選択肢  $a$  を所与として、ステップ 3 で構築した誤答選択肢作成方

表 1 誤答選択肢作成方略

ラベル	方略	方略の説明
1	正反対の事実の利用	正答選択肢と正反対の事実に言及した誤答選択肢を作成
2	無関係な事実の利用	本文中で一切言及されていない内容に関する誤答選択肢を作成
3	事実の不正な組み合わせ	本文中の事実を不適切に組み合わせた誤答選択肢を作成

表 2 誤答選択肢生成時に Llama3 に与える指示プロンプト

Here is a strategy to generate high-quality distractors: { 誤答選択肢作成方略 $y$ の説明 }
You are an expert in creating multiple-choice question for reading comprehension. You are provided with a set consisting of a reading passage, a question, a correct answer. The question and correct answer are all related to the content of the reading passage. Your task is to create a distractor (an incorrect option) from reading passage, question, and answer according to the above strategy. Please follow the following steps: 1. Firstly, understand the above strategy well. 2. Carefully read and understand the provided reading passage, question, and correct answer. 3. Following the strategy, generate a distractor considering the given reading passage, question, and correct answer. 4. Output the distractor you generated ONLY. The input passage, question, and correct answer will be given the following format: ### Reading Passage: $c$ ### Question: $q$ ### Answer: $a$
{3-shot samples; $j \in \{1, 2, 3\}$ の提示 }
### Reading Passage: $c_j^{(top3)}$
### Question: $q_j^{(top3)}$
### Answer: $a_j^{(top3)}$
### Distractor: $d_j^{(top3)}$

略選定機によって適切な方略を選定する。次に、選定された方略  $y$  と読解文  $c$ , 問題  $q$ , 正答選択肢  $a$  を所与として、表 2 のプロンプトを Llama3 に入力して誤答選択肢を生成させる。なお、プロンプト内では、指定した方略に関連する few-shot サンプルの提示も行う。この際、few-shot サンプルは入力された読解文  $c$ , 問題  $q$ , 正答選択肢  $a$  との関連性が高いデータを  $\mathcal{D}_{train}$  から次の手順で選択して与える。(1)  $c$ ,  $q$ ,  $a$  を連結したものを SimCSE (Simple Contrastive Learning of Sentence Embeddings) モデルに入力し、分散表現ベクトルを取得する。(2) 同様に、 $\mathcal{D}_{train}$  中の全てのデータに対しても分散表現ベクトルを計算しておく。(3)  $\mathcal{D}_{train}$  の中から、指定した方略  $y$  に対応しており、かつ、入力に対応する分散表現ベクトルとのコサイン類似度が高いデータを 3 つ選択し、それを few-shot サンプルとしてプロンプトに加える。表 2 では、選ばれた 3 つの few-shot サンプルを  $\left\{ \left( c_j^{(top3)}, q_j^{(top3)}, a_j^{(top3)}, d_j^{(top3)} \right) \right\}_{j=1}^3$  と記述する。

### 3 評価実験

評価実験ではデータセット  $\mathcal{D}$  として RACE データセットを利用し、提案手法のステップ 2 と 3 を行った上で、 $\mathcal{D}_{test}$  の各データに対してステップ 4 に従って誤答選択肢を生成した。本実験では、提案手法との比較のために、以下の 2 つの手法でも誤答選択肢を生成した。

比較手法 1: 提案手法に誤答選択肢作成方略と few-shot サンプルのどちらも与えずに誤答選択肢を

表 3 生成された誤答選択肢の品質評価

評価観点	提案手法	比較手法 1	比較手法 2
A	<b>0.97</b>	0.93	0.95
B	<b>0.56</b>	0.20	0.32

生成する手法

比較手法 2: 提案手法に基づくが、few-shot サンプルのみをランダムに与えて誤答選択肢を生成する手法

各手法で生成された誤答選択肢の品質は、実際に誤った選択肢であるか (評価観点 A) と、簡単には排除できない選択肢か (評価観点 B) の 2 つの観点に基づいて人手により評価した。評価観点 A については 2 段階 (0: 不適切, 1: 適切) で評価し、評価観点 B については 0: 迷わしとして機能しておらず、容易に誤答選択肢であると判断できる, 1: 迷わしとして機能しており、読解文を適切に理解することで誤答選択肢であると判断できる, の 2 段階で評価した。

実験結果を表 3 に示す。表の数値は 100 問に対する評価点の平均値を示す。ここで、手法間での平均性能に有意な差があるかを確認するために、評価観点ごとに分散分析を行ったところ、評価観点 B において、有意水準 1% で有意差が見られた。そこでさらに、評価観点 B について多重比較を行ったところ、提案手法と比較手法 1、提案手法と比較手法 2 の間で有意水準 1% で有意差が見られた。以上の結果から、誤答選択肢作成方略を指定した上で few-shot サンプルを活用する提案手法がより高品質な誤答選択肢を生成できることが確認できた。

### 4 まとめ

本研究では、多肢選択式読解問題の誤答選択肢の生成に焦点を置いて、品質の高い誤答選択肢の生成を目指す手法を提案した。

#### 参考文献

- (1) Dutulescu, A., Ruseti, S., Iorga, D., Dascalu, M., McNamara, D.S.: “Beyond the obvious multi-choice options: Introducing a toolkit for distractor generation enhanced with NLI filtering,” Proceedings of the 25th International Conference on Artificial Intelligence in Education, pp.242-250, 2024
- (2) Maity, S., Deroy, A., Sarkar, S.: “A Novel Multistage Prompting Approach for Language Agnostic MCQ Generation using GPT,” Proceedings of the 46th European Conference on Information Retrieval, pp.268-277, 2024