

# 英単語穴埋め問題における選択肢の自動生成に向けて

吉見菜那, 廣中勇希, 梶原智之 (愛媛大学), 荒瀬由紀 (大阪大学), 内田諭 (九州大学), 二宮崇 (愛媛大学)

## 1. 概要

- 英単語穴埋め問題
  - 空欄を持つ問題文, 選択肢 (正解1つ, 不正解3つ) で構成

Jeff didn't accept the job offer because of the \_\_\_\_ salary.  
(a) low (b) weak (c) cheap (d) inexpensive

- 英単語穴埋め問題の不正解選択肢を自動生成するタスク
  - 選択肢の自動生成によって教師の負担を軽減できる
- 提案手法では, 問題ごとに候補とする順位の閾値を変更し, 妥当性と信頼性の両方を改善した

## 2. 先行研究: 類似単語100件の生成 → BERTによるリランキング



- 単語埋め込み + BERTリランキング<sup>[1]</sup>
  - 正解単語に意味が近い単語を生成
  - BERTリランキング
  - 11-39位を候補とする
- 折り返し翻訳 + BERTリランキング<sup>[2]</sup>
  - 英語 ⇄ ピボット言語に折り返し翻訳後, 単語アライメントを計算
  - BERTリランキング
  - WordNetによる正解単語の同義語を除き, 同じ品詞のみを候補とする

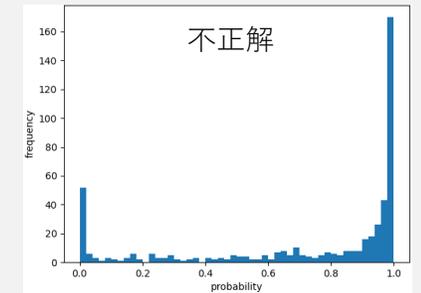
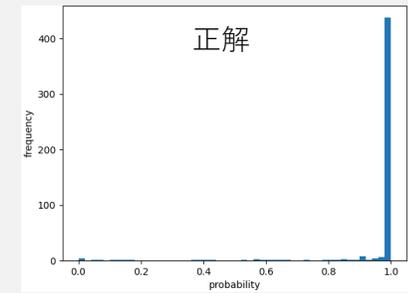
既存の手法では取得する候補の順位が固定されている

[1] Yeung+ (2019) Difficulty-aware Distractor Generation for Gap-Fill Items  
[2] Panda+ (2022) Automatic Generation of Distractors for Fill-in-the-Blank Exercises with Round-Trip Neural Machine Translation

## 3. 正解選択肢と不正解選択肢のBERT予測確率の分析

BERTの穴埋めによるスコア

- 正解単語の予測確率
  - 平均値: 0.958
  - 分散: 0.025
- 不正解単語の予測確率
  - 平均値: 0.724
  - 分散: 0.125



不正解単語の予測確率は, 正解単語の予測確率より低く, ばらつきがある

## 4. 提案手法: BERT予測確率に基づく動的なフィルタリング

問題ごとに, 不正解選択肢の候補とする順位の閾値を変更

$$\text{不正解単語の予測確率} \leq (\text{正解単語の予測確率} \times \text{任意の値 } \lambda)$$

## 5. 評価実験: 大学入試の問題500件 (2017-2021年)

妥当性(Recall@3): 実際の不正解選択肢を生成できた割合

信頼性(3-gram): 候補を含む問題文の3-gramがWikipediaに現れない確率

候補生成 + リランキング	フィルタリング	MRR	妥当性(Recall@3)	信頼性(3-gram)
単語埋め込み(fastText) + BERTリランキング	11-39位 <sup>[1]</sup>	0.169	3.5%	79%
	$\lambda = 0.97$	0.176	4.1%	<b>83%</b>
	$\lambda = 0.98$	0.187	4.7%	81%
	$\lambda = 0.99$	<b>0.209</b>	<b>5.4%</b>	80%
折り返し翻訳 + BERTリランキング	WordNet + POS <sup>[2]</sup>	0.100	8.5%	72%
	$\lambda = 0.97$	0.111	8.3%	<b>79%</b>
	$\lambda = 0.98$	0.131	8.6%	78%
	$\lambda = 0.99$	<b>0.142</b>	<b>9.8%</b>	77%

提案手法は, 妥当性と信頼性の両方を改善した