

2025 年度 高等学院同窓会学術研究奨励金
研究成果報告書概要（WEB 公開用）高等学院 長
高等学院同窓会理事長 殿

研究代表者氏名 [楠山大貴]

学年・組・番号 [3 年 G 組 38 番]

研究課題： 推論時スケーリング則に意図推定能力の向上は従うのか

(英文) Do improvements in intent estimation ability follow scaling laws during inference

研究概要：

(研究課題を選んだ動機、達成するための計画・目的・方法等について 200～400 字で記入してください)

ChatGPT 等の大規模言語モデル (LLM) において、モデル規模や計算量に伴い性能が向上する「推論時スケーリング則」は広く知られているが、ユーザーの不足情報を補う「意図推定能力」がこの法則に従うかは未解明である。本研究は、対話システムに不可欠なこの能力を定量的に検証することを目的とする。具体的には、質問生成能力を測る IN3 データセットを用い、高性能な GPT-5 と軽量の Gemini 2.5 Flash の性能差を比較する。情報の不足度をタスク難易度と定義し、難易度が上がるほどモデル間の性能差が拡大するという仮説を立て、高難易度データとそうではないデータを用いて検証を行う。実験の結果、仮説に反して性能差は難易度によらず一定であり、意図推定は従来の推論深度とは異なる能力軸である可能性が示唆された。本研究は、推論時スケーリング則の適用範囲を再検討し、次世代の AI 評価における新たな視点を提供するものである。

研究成果：

(研究の結果概要、結果に対するフィードバックや感想等について 200～400 字で記入してください)

実験の結果、GPT-5 と Gemini 2.5 Flash の質問包含率の差は、全データセットにおいて約 23% でほぼ一定であった。難易度が高いほど性能差が拡大するという仮説に反し、意図推定能力はモデルの規模や推論深度に依存する「推論時スケーリング則」に必ずしも従わないことが明らかになった。また、高難易度タスクにおいてスコアがわずかに上昇する「タスク強制効果」を確認した。これは、情報の曖昧さがモデルに質問を強いる結果、必須質問との一致度を高めたものと推察される。仮定とは異なった結果となったが、有意義な結果にはなったと思っている。

研究者：(以下の、代表者・分担者は学年・組・氏名を明記する)

研究代表者 楠山 大貴

研究分担者

担当教諭 八百幸大 先生

(受給額：25000円)

※研究課題、研究概要、研究成果、研究代表者名が WEB ページ上で公開されることに同意します
(次のページに続きます)

研究成果写真：
(研究過程がわかる写真や、研究結果がわかる写真などを数点貼り付けてください)

卒論_検証_2号.ipynb

★

ファイル

編集

表示

挿入

ランタイム

ツール

ヘルプ

コマンド

+ コード

+ テキスト

すべてのセルを実行

接続

ランダム

gpt

↑

↓

↺

↻

⋮

```

import json
import os
import time
from openai import OpenAI
from openai import APIError
import random
from concurrent.futures import ThreadPoolExecutor, as_completed

# --- 1. JSONファイルの読み込み ---
input_file_path = '/content/formatted_data_ia_sampled.json'
try:
    with open(input_file_path, 'r', encoding='utf-8') as f:
        tasks_data = json.load(f)
    tasks = tasks_data
    print(f"[✓] {input_file_path}から{len(tasks)}個のタスクを正常に読み込みました。")
except FileNotFoundError:
    print(f"[✗] エラー: ファイルが見つかりません - {input_file_path}")
    exit()
except json.JSONDecodeError:
    print(f"[✗] エラー: JSONファイルの形式が正しくありません - {input_file_path}")
    exit()
except Exception as e:

```

変数

ターミナル

ファイル(F)

編集(E)

表示(V)

Git(G)

プロジェクト(P)

デバッグ(D)

テスト(S)

分析(N)

ツール(T)

拡張機能(X)

ウィンドウ(W)

ヘルプ(H)

検索(Ctrl+Q)

ソリューション

サインイン

Live Share

100%

問題は見つかりませんでした

in3_upstream_ja (t).json

スキーマ: スキーマが選択されていません

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44

```

100%

問題は見つかりませんでした

in3_gpt.task_outputs.json

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

```

以上