

進捗報告

1 やったこと

- モデルマージ動作確認
- 7b のモデルにおけるファインチューニング

2 モデルマージ動作

先週 7b 程度のモデル 2 つであればローカルのサーバーで進化的モデルマージを試すことができることを確認した。他にどの組み合わせが実行可能なのかを検証した。結果として, Llama3-8b と Gemma2-9b のようなモデルのアーキテクチャが異なる場合, Gemma2 のモデルのアーキテクチャの Gemma2ForCausalLM のようにライブラリが対応していないモデルに関しては実行不可能だった。

モデルのアーキテクチャを揃えていろいろ試した中で, 実行可能だった組み合わせとしては以下の通り。

- elyza/ELYZA-japanese-Llama-2-7b-instruct¹ × stabilityai/japanese-stablelm-instruct-beta-7b²
どちらも Llama-2 を基盤としているモデル。
- stabilityai/japanese-stablelm-instruct-gamma-7b³ × augmxnt/shisa-gamma-7b-v1⁴
どちらも Mistral-7B-v0.1 を基盤としている。

3 7b のモデルのファインチューニング

7b で安定した出力をするモデルを探すため, 前節で述べた 4 つのモデルのファインチューニングをした。

3.1 elyza/ELYZA-japanese-Llama-2-7b-instruct

基本的に 2 文で応答することが多く, 会話の応答としては少し長いかもしれないが十分に自然な出力ができている。

3.2 stabilityai/japanese-stablelm-instruct-beta-7b

ファインチューニング後の出力は複数文が出力され安定した出力とはならなかった。

モデルのファインチューニングの際に明示的に加えた EOS トークンが出力の中に見られた。なぜ EOS トークンが出てきているのに出力がそれ以降も続いているかは確認の余地があるが, プログラム側で EOS トークンが出現するまでを LLM の出力としてあげることで自然な応答を得られる可能性が高い。

また, ファインチューニング前の LLM の出力持たのモデルと比べて性能の低さが感じられる。

¹<https://huggingface.co/elyza/ELYZA-japanese-Llama-2-7b-instruct>

²<https://huggingface.co/stabilityai/japanese-stablelm-instruct-beta-7b>

³<https://huggingface.co/stabilityai/japanese-stablelm-instruct-gamma-7b>

⁴<https://huggingface.co/augmxnt/shisa-gamma-7b-v1>

3.3 stabilityai/japanese-stablelm-instruct-gamma-7b

返答がかなり長い文になっており、出力トークンの上限を 64 としているが上限に届いたため文の途中で不自然に途切れている出力が多く見られた。

3.4 augmxnt/shisa-gamma-7b-v1

stabilityai/japanese-stablelm-instruct-gamma-7b と似たように出力トークンの上限に達して不自然に文が途切れるような出力が多く見られた。

4 結果

Mistral-7B-v0.1 を基盤としている stabilityai/japanese-stablelm-instruct-gamma-7b, augmxnt/shisa-gamma-7b-v1 が似たような傾向を示した。

Llama2 を基盤としている elyza/ELYZA-japanese-Llama-2-7b-instruct, stabilityai/japanese-stablelm-instruct-beta-7b は前者が応答として自然な応答をしているが、後者は不安定な出力をしている。基盤モデルの影響が出力の安定性に大きく関わっていると考えていたが、事後学習の違いもファインチューニング語の性能に影響を及ぼしている可能性が考えられる。

マージとして使えるレベルの出力が得られているモデルとしては Llama2 派生の 2 つなので、stabilityai/japanese-stablelm-instruct-beta-7b の出力の加工をある程度プログラム側で制御することでマージの際の評価をしていく。

5 今後やること

- ロールプレイをするタスクにおけるモデルマージ (最優先)

定量評価用の独自タスクの実装とマージしたモデルの評価までやりたかったが、間に合わなかった。

後期研究会では結果はどうであれモデルマージまで行きたいので急いでやります。

参考文献