

進捗報告

```
>> User:Hello
DialoGPT: Hiya
>> User:What your name?
DialoGPT: I'm a girl
>> User:How is the weather today?
DialoGPT: It's nice
>> User:It sounds good
DialoGPT: I'm glad
>> User:Good bye
DialoGPT: I'll see you around
```

図 1: 対話例

1 今週やったこと

- 対話タスクの動作確認

2 英語での対話

図 1 に対話例を示した。英語の場合、対話を自然に続けられることを確認した。

3 日本語での対話

Tokenizer がテキストを前処理し、Model が出力するという関係になっている。Tokenizer にエラーが起きて、動かすことができなかった。

GPT に準拠するなら BPE(Byte Pair Encoding) によって単語を分解し Token に置き換えるが、日本語には文字の種類が多いことや空白がないことなどが原因で適切な Tokenizer を英語のものとは変える必要がある。BPE の他に、形態素解析 (Mecab) を元にした BERT の Tokenizer もある。学習の前提となるので、学習済みパラメータがどのような Tokenizer を使用したのかを見なければならぬが、github の README を見ても情報が少なく現在は動作を確認する段階に至らなかった。

4 今後の予定

- 日本語での対話タスクの動作確認