README

英法机器翻译

对英法平行语料进行适当的预处理，并使用 Pytorch 深度学习框架构建 Transformer 模型并在英法平行语料上进行训练

本次实验使用的数据集是英法平行语料库，包含3个文件：english.txt和french.txt为只含有英语和法语句子的语料，需要从中统计出英语和法语的词汇表，fra.txt中则提供了英语-法语句子的对照，需要自行分割为训练集（90%）和验证集（10%）

使用 Pytorch 构建 Transformer 模型，具体实现可以使用torch.nn.Transformer（简单易用，但相对更加难以理解），也可以参考知名开源工具包 https://github.com/huggingface/transformers （更好理解），如果模型核心算法部分使用开源代码则需对源代码添加逐行注释

使用构建好的 Transformer 模型在训练集上进行训练，训练细节可以参考 https://arxiv.org/pdf/1706.03762.pdf

在验证集上对模型进行评估，使用 BLEU 值作为评估指标

数据预处理

运行`data/preprocess.py`得到预处理之后的数据`data/processed`

模型及评估指标

- 阅读`model/transformer.py`并补充完成multihead attention
- 补充完成`./metrics.py`中BLEU的计算

模型训练

- 运行`./main.py`