

域适应翻译

准备资料

预训练模型

Facebook开源的M2M-100，参
数约1.9GB大小

开源库

huggingface Transformers

微调数据集

已有翻译好的领域内平行语料库
(已对齐)

领域内单语数据，经过翻译引擎
翻译后，作为微调数据

翻译术语字典

领域内的专业翻译术语字典

微调训练

微调时使用fp16和gradient
checkpoint来减少GPU的显存使
用量

BLEU 38.5左右

Beam search 或topk寻找最好的
翻译结果

预测阶段

使用微调后的模型预测，并且首
先对自定义的专业术语词库使用
模型进行翻译，然后在模型翻译
后对其原文和翻译后文本进行匹
配，如果匹配，那么就进行术语
替换。

对比后的翻译结果好于Google翻
译

预测时动态加载模型，5分钟不
预测，就动态卸载模型

任务Flask接口，把给定PPT翻译
成英文，支持自定义excel术语字
典，支持默认从mysql读取翻译
术语字典

其它

普通领域的翻译到化妆品领域的
翻译模型