本周进度

1. 使用古诗词语料对 GPT - 2 模型进行训练；
2. 看 GPT - 2 模型源码 -- `部分完成`；
3. 看 NeteaseCloudMusic 代码 -- 爬取网易云音乐数据 -- `部分完成`；
4. 看论文 -- `部分完成`。

1. GPT - 2 模型训练

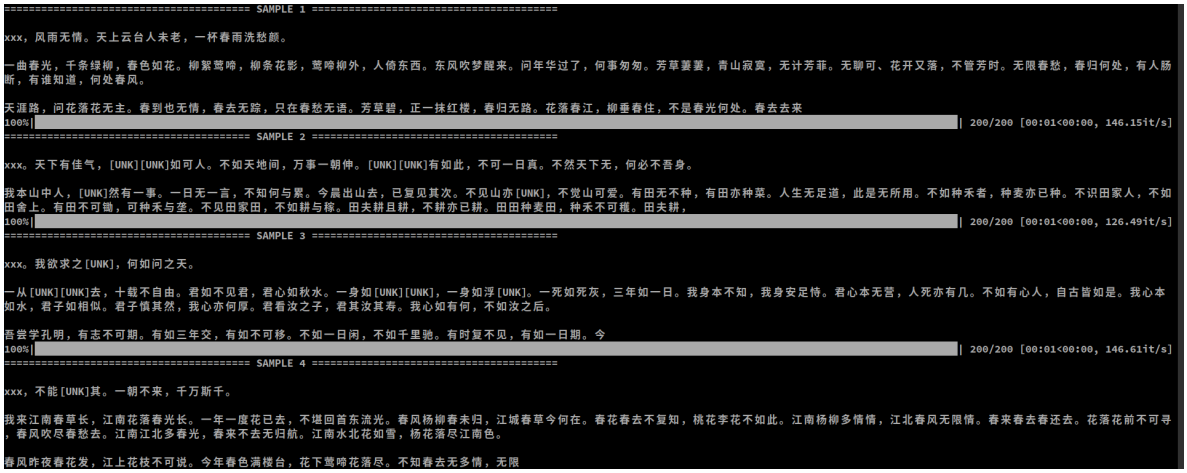
1.1 古诗词部分

1.2.1 语料

- 古诗词集
- 训练集大小： 共85w+ 行， 187M
    - 预处理

1.2.2 训练

- 预训练 -- `使用默认循环次数-5`
    - **同小说预训练**
- 生成结果



2. 问题总结

- 爬取网易云音乐数据所使用项目可能爬取不成功 -- `还没来得及实践`
    - 项目未给出使用方式，需要看一下源代码；
    - 项目两年前更新，可能会爬取不成功

3. 本周任务

- 总结 GPT - 2；
- 实践测试爬取网易云音乐数据 -- 尝试查找一些其他爬虫方法；
- 看论文。