参考资料

Wenet的尝试冲冲冲

体验了开源的离线语音识别模型 wenetspeech，准确度很高。分享一些使用心得。 - 糯米PHP

TIMELINE：

- [ ] 0124-0130：完成相关技术的写作
- [ ] 0204-0211：研究实时语音识别怎么做
  - [ ] paddle尝试：模拟语音字节流
  - [ ] wenet尝试：直接使用他们的模型来做
  - [ ] 研究怎么在不同的端显示收到的文字流---不行就自己做demo
- [ ] 0212-0228：完成论文剩下的部分

卷积神经网络(CNN)在语音识别中的应用_园荐_博客园

https://blog.csdn.net/l641208111/article/details/102496061

https://zhuanlan.zhihu.com/p/123211148

https://zybuluo.com/hanbingtao/note/541458

EM算法详解

webrtc VAD 算法_liwenlong_only的博客-CSDN博客_vad算法

WebRTC VAD算法初探 - 程序员大本营

https://zhuanlan.zhihu.com/p/387581644

webrtc VAD 算法_liwenlong_only的博客-CSDN博客_vad算法

https://blog.csdn.net/shichaog/article/details/52399354/

浅谈MFCC_fengzhonghen的专栏-CSDN博客_mfcc

0110-0115

- [x] ~~完成毕业论文的整体思维导图~~
- [x] ~~完成标点加入的代码开发，打包新的docker，简化代码~~
- [x] ~~在服务器上测试新的docker代码是否可用~~

语音识别方案 - 联想语音团队多场景实时语音文字转换方案_beijing_xsw的博客-CSDN博客

远场语音识别面临的瓶颈与挑战

做完以上后和老师确认是否ok，和彪哥同步进展

- [x] ~~解决没有标点符号的问题：这个可以用paddlespeech现成的ernie模型来做，但是会影响多进程的正常使用~~

python调用讯飞语音听写(流式版)_sinat_41787040的博客-CSDN博客_python调用讯飞语音

0110-0114

- [ ] 完成qtav的编译及简单播放器的开发
- [ ] 和老师确认一下目前的开发情况是否可以完成一篇毕业论文，如果不行还差哪里
- [ ] 找王老师要一下学堂的代码想办法集成进去
- [x] ~~继续研究实时语音识别，python实时传输音频流怎么弄（再看看wenet能不能拿来使）：放弃研究~~

0110-0105

https://github.com/wang-bin/QtAV

https://blog.csdn.net/qq_38766431/article/details/115328784

- [ ] 周三：
  - [x] ~~研究实时音频识别的可能性，看怎么能够接入智慧教室~~

【python简短实用的小代码】:利用电脑麦克风录制一段音频_pikapika_chu的博客-CSDN博客

https://blog.csdn.net/weixin_31674039/article/details/113499219

- [x] ~~找张荆要一下心神课堂的学生端客户端代码，进行播放器改造~~

【C++】使用 libass，完成 Direct3D 11 下的字幕渲染 - 最后的绅士 - 博客园x

新的镜像：docker pull paddlepaddle/paddle:2.2.1

目前面临的问题：docker里无法直接pip install paddlepaddle，必须自己拉image来编译

再研究一下实时语音识别的可能性：

1. 写测试程序，模拟发送字节流给实时预测程序

关于docker的尝试：

1. 编写测试程序--一个py文件，功能为读取指定目录下的mp4文件，并生成对应的mp4名称的txt文件到同级目录，需要传入参数（参数为指定目录）
2. 编写dockerfile
   a. 启动docker时候怎么向内部程序传入参数 - tianyee - 博客园
   b. Docker实战-编写Dockerfile_DWWWWWEI的博客-CSDN博客_dockerfile编写
3. 将python程序打包为docker镜像
4. Run docker，设置数据卷，将本地需要操作的目录挂载到docker中，并传递参数

```
1 docker run -v {host_mount_dir}:{docker_mount_dir} -e PARAMS= {--monitor_path=doc
```

5. 关闭docker

👀如果可行的话，自己的程序的修改方式为：

1. 从docker传入监控目录的参数

```
1 --monitor_path={docker volume path}
```

2. 修改日志存储的位置，日志存储在监控目录的同级目录下（判断是否有logging目录）done
3. 给docker传递的参数就是volume挂载在容器中的位置
4. 整个主进程如果异常退出需要重启，重启机制

Docker数据卷挂载命令volume(-v)与mount的总结_Charles Shih 技术博客-CSDN博客_docker mount

如何将python应用制作成容器镜像? - 知乎

现在的问题：

在local.py里开了监控文件和语音识别两个进程，但是local里加载的模型和其他进程是独立的，导致每次识别都要重新加载一次模型--已解决

多进程解决python处理慢的问题

使用docker来做分布式：使用Docker时如何将文件写入主机?-python黑洞网

---

logger重复打印的解决方法：Python | 解决方案 | 多个文件共用logger，重复打印问题_小天使faith的专栏-CSDN博客---已解决

改成原来的argument的形式来print数据---已解决

遇到的问题

1. 怎么样把路径取出来---已解决，用队列来解决
2. 在监控这个目录的时候，可能随时有新增的文件，怎么保证每一个新增文件都被语音识别并且生成了字幕---已解决，使用队列来解决

queue --- 一个同步的队列类 - Python 3.10.1 文档

多进程共享队列数据：Python 进程，进程间通过队列共享数据，队列Queue_houyanhua1的专栏-CSDN博客_python 进程共享queue

1227-1231

现在的方案换了，一周内完成这个方案：

1. 弃用http服务，改成读指定路径做监控，只处理mp4文件---done（只有mp4被放进队列）
2. 5分钟起一次语音识别服务，考虑容错--如果异常关闭需要重新启动
3. 把字幕文件放在视频文件的相同位置
4. 字幕文件支持两个格式srt和vtt格式
5. 分布式部署--尝试用docker来部署（可以咨询下磊哥）
6. 加上只处理中文课程的逻辑（但是论文中其实可以不体现这个逻辑）
7. 还要设计一个通用的日志类

Web 视频文本轨格式（WebVTT） - Web API 接口参考 | MDN

Python Watchdog--监控文件系统事件_XerCis的博客-CSDN博客_python watchdog

一个更加清楚一点的教程：Gunicorn及Nginx的配置 - 本人小白 - 博客园

浏览器js视频播放器外挂srt字幕(vue为例)_Zyrzka的博客-CSDN博客_js解析srt文件

使用**多线程**技术将一台机器的能量最大化

Linux安装nginx - star-xin - 博客园（暂时还没做防火墙哪一步）

Gunicorn小白入门简介

nginx负载均衡的5种策略及原理_zxing的博客-CSDN博客_nginx负载均衡策略

gunicorn实现flask并发_wuyongpeng的专栏-CSDN博客_gunicorn 协程

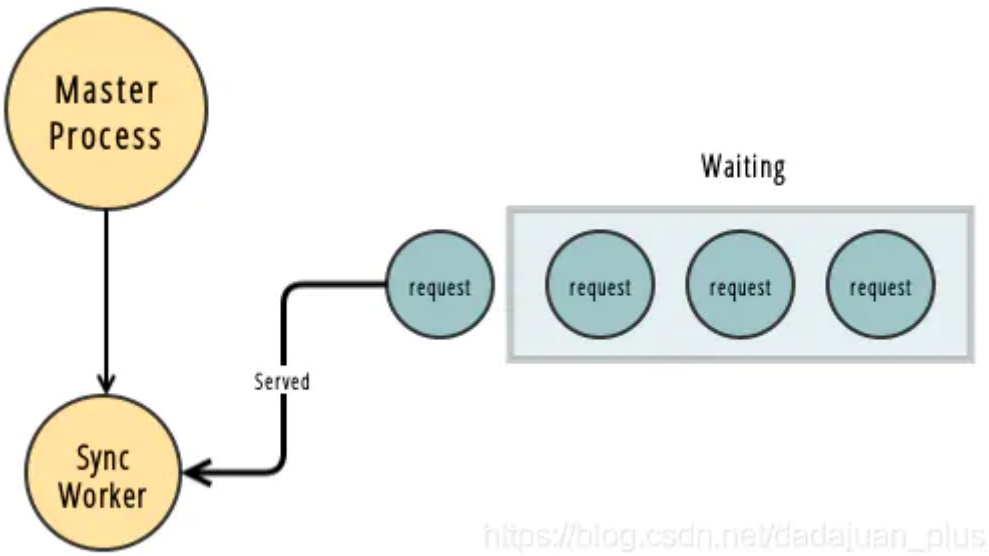
测试机器上：`/usr/local/python`是存放python3.7的位置，beamsearch需要用3.7以上的版本

python3.6在`/usr/local/bin/python3.6`里

飞桨PaddlePaddle-源于产业实践的开源深度学习平台

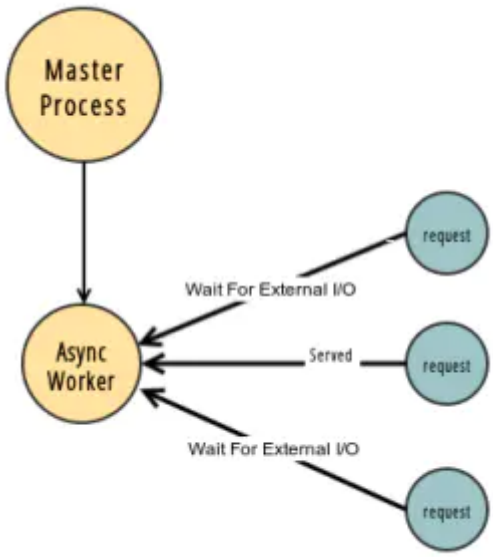
Gunicorn学习笔记

Sync Workers



每个worker进程，一次处理一个请求，如果有其他请求被分配到该worker进程，则阻塞，需要第一个请求完成，一个请求一个进程。**适合访问量不大，CPU密集而不是IO密集的情形**。

Async Workers



基于Greenlet实现：微线程

- Gevent：实现异步模型，可以自动切换，保证总有微线程在运行，无需等待IO

以下内容在周二前完成

- [x] ~~和老师确定最终的产品形态：已经发微信询问了，让我自己想~~
- [ ] 先在本机测试完成gunicorn+flask+nginx的服务部署
- [x] ~~完成wav文件缓存删除的逻辑~~
- [x] ~~日志模块：还需要完善~~

以下内容及需求需要和彪哥及PM确认--周三前完成确认

1. 如果是直接把字幕压制进视频的话
- [ ] 完成相应的python代码，注意要处理不翻译英语课的逻辑
- [ ] 跑通整个流程

2. 如果不是直接把字幕压制进视频的话
- [ ] 完成客户端是否开启字幕的开发：
  - [ ] 跳过英语课--英语课暂不支持字幕
  - [ ] 拿到课程对应的classid，用classid关联字幕
  - [ ] 字幕组件以及组件和屏幕变换时的适配
- [ ] 完成网页端是否开启字幕的开发：
  - [ ] 跳过英语课--英语课暂不支持字幕
  - [ ] 拿到课程对应的classid，用classid关联字幕
  - [ ] 字幕组件的开发：需要js来做
- [ ] 完成移动端是否开启字幕的开发：
  - [ ] 跳过英语课--英语课暂不支持字幕
  - [ ] 拿到课程对应的classid，用classid关联字幕
  - [ ] 字幕组件的开发：需要java来做

3. 实时字幕的尝试：等先把离线的做完再研究实时的咋做吧
- [ ] 客户端推流一份到服务器去
- [ ] 语音识别服务返回文字
- [ ] 在不同的端展示文字--集群转发过去展示

1213-1217

深度学习多线程部署-学习笔记_studyeboy的专栏-CSDN博客_深度学习多线程

使用Gunicorn部署Flask Web服务

qt里写播放器：GitHub - hqy2000/CaptionPlayer: 双屏幕双语字幕播放器。

1213要做的事情

- [x] ~~完成噪声处理（加窗的方式）~~
- [ ] 完成wav文件缓存删除的部分
- [x] ~~把整个流程的python代码写完~~

机器学习模型持续部署(基于Flask, Docker, Jenkins 和 Kubernets)_Mr_不想起床的博客-CSDN博客

python三大神器之virtualenv - 似是故人来~ - 博客园

1206-1210

python3.7生成字幕文件_洪源的博客-CSDN博客_python生成字幕文件

登录不上github的解决办法

立体声转单声道ffmpeg

ffmpeg转采样率

Python调用百度API实现语音识别(二)_咪哥杂谈-CSDN博客

librosa重采样

flask服务部署--应用可以持续运行

python -- gunicore 调用 flask 详细配置_justlpf的专栏-CSDN博客

音频文件按照正常语句，断句拆分的处理方法_watfe的专栏-CSDN博客_语音断句

Python pydub实现语音停顿切分_wangqianqianya的博客-CSDN博客_pydub.silence

MILESTONE

- [x] ~~Python Flask服务搭起来：参考paddlepaddle的flask实现~~
- [x] ~~完成FLV转wav的脚本：使用moviepy就可以做到，还可以设置采样率~~
- [x] ~~完成ACC转wav的脚本：https://www.cnblogs.com/shiyublog/p/11245791.html#_label2~~
- [x] ~~长音频处理的脚本：参考paddlepaddle的webrtcvad实现~~
- [x] ~~能够生成字幕文件（明确时间轴的脚本：https://www.cnblogs.com/tocy/p/subtitle-format-srt.html~~
- [ ] 可以参考的语音项目：https://github.com/zhoudd1/python-Speech_Recognition

- [ ] PCM数据转化为wav格式
- [ ] https://bbs.sangfor.com.cn/forum.php?mod=forumdisplay&fid=49
- [ ] python创建历史文件：https://zhuanlan.zhihu.com/p/359751690

1206

4. 需要把声音文件从立体声转成单声道：Python将立体声转为单声道：不可用，会有音质损失

1129-1203

用到的技术MinGW：https://blog.csdn.net/YahamaTarGe/article/details/89380164

https://my.oschina.net/u/2501904/blog/1162753

sanic部署：https://blog.csdn.net/cdknight_happy/article/details/110373458

RTMP拉流：https://blog.csdn.net/Hanghang_/article/details/104813126?spm=1001.2101.3001.6650.2&utm_medium=distribute.pc_relevant.none-task-blog-2%7Edefault%7ECTRLIST%7Edefault-2.no_search_link&depth_1-utm_source=distribute.pc_relevant.none-task-blog-2%7Edefault%7ECTRLIST%7Edefault-2.no_search_link

两周内尝试的新路：使用Pytorch来做，先不尝试用paddlepaddle

- [x] ~~在测试机上搭好一个输入长wav文件可以返回其对应时间及文本的服务~~
- [x] ~~看完XSStream部分的代码，明确在哪部分加入服务访问并能够把字幕加上去（感觉像是要在后端做了，而不是在客户端做）~~
  - [x] ~~客户端做的方法：在服务器上存储视频和其字幕文件的对应关系，每次客户端放视频只是读取对应关系来播放（需要建数据库，可以在客户端实现数据库的读写和创立，用MongoDB来做）~~
  - [x] ~~服务端做的方法：在服务器上重新压制视频，这样分发下去的视频都是一样的~~

MILESTONE

- [x] ~~Python Flask服务搭起来：参考paddlepaddle的flask实现~~
- [ ] 完成FLV转wav的脚本
- [ ] 完成ACC转wav的脚本：https://www.cnblogs.com/shiyublog/p/11245791.html#_label2
- [x] ~~长音频处理的脚本：参考paddlepaddle的webrtcvad实现~~
- [ ] 能够生成字幕文件--明确时间轴的脚本：https://www.cnblogs.com/tocy/p/subtitle-format-srt.html
- [ ] 可以参考的语音项目：https://github.com/zhoudd1/python-Speech_Recognition

- [ ] PCM数据转化为wav格式
- [ ] https://bbs.sangfor.com.cn/forum.php?mod=forumdisplay&fid=49
- [ ] python创建历史文件：https://zhuanlan.zhihu.com/p/359751690

1122-1126

https://www.jb51.net/article/207661.htm VS2019添加第三方库

如果使用paddlepaddle必须要走的路：

5. C++推理-编译paddlepaddle的windows推理库
6. 将模型的inference脚本、数据处理部分都用C++重新写一遍（粗略看了一下，超级恼火）
7. 说服老师在线推理（但是实验室没有服务器这些的，需要了解后端改动机制）

cmake教程：https://blog.csdn.net/yaoyuanyylyy/article/details/79024962

解决cmake-gui编译失败的问题：

- https://www.cnblogs.com/qq2806933146xiaobai/p/13359446.html
- https://blog.csdn.net/diaodaa/article/details/106122943

再换个思路：把所有的python都整成dll来搞

走一条模型上的新路：使用kaldi来做（好像更难，一个kaldi服务器）

x86window部署：https://paddleinference.paddlepaddle.org.cn/demo_tutorial/x86_windows_demo.html

pytorch编译为C++：https://oldpan.me/archives/pytorch-c-libtorch-inference

https://github.com/cpuimage/WebRTC_VAD webrtc-vad实现

一个新的方式：https://github.com/wenet-e2e/wenet#installation 使用wenet来做语音识别

wenet的介绍：https://www.cnblogs.com/dahu-daqing/p/14805678.html

https://paddle-inference.readthedocs.io/en/latest/demo_tutorial/x86_windows_demo.html

编译PaddlePaddle的预测库

CMakeList的位置：https://github.com/PaddlePaddle/Paddle-Inference-Demo/blob/master/c%2B%2B/lib/CMakeLists.txt

VS2019 CMake的小办法：https://www.jb51.net/article/180463.htm

https://blog.csdn.net/a2824256/article/details/117806304

https://blog.csdn.net/a2824256/article/details/118112448

编译resnet作为一个小小的尝试：https://github.com/PaddlePaddle/Paddle-Inference-Demo/tree/master/c%2B%2B/resnet50

1115-1119

- [x] ~~使用好未来的数据集完成训练（DeepSpeech）~~
- [x] ~~理清楚DeepSpeech的代码逻辑~~
- [x] ~~完成windows上的paddle-lite的编译~~

接下来要做的事情：

模型保存的常见问题

模型保存及加载

路径1:windows上的GPU预测部署示例➡官方教程

~~路径2:pdparams转化为ONNX➡官方教程：行不通，不支持load算子~~

路径3:先编译Paddle-lite再把模型导出为对应的格式➡官方教程1 官方教程2

路径4:直接使用C++ API来做

路径5:飞桨推理产品介绍

windows预测库介绍

C++预测示例

VS2019的预测库编译 （推荐的step1！）

1101-1105

- [ ] 看完MARS的项目代码，理清楚逻辑
- [x] ~~windows配置deepspeech的环境尝试跑起来中文模型~~
- [ ] 看完deepspeech的英文论文

关键参考资料：

8. windows系统下安装anaconda：

   https://www.jianshu.com/p/a0e386427944（用conda做虚拟环境）

   https://blog.csdn.net/honest_boy/article/details/89600042（主要参考）

   https://blog.csdn.net/weixin_44776894/article/details/106159483

   https://blog.csdn.net/LaughingMei/article/details/109742489

9. deepspeech实践1：pytorch下使用deepspeech2

   https://github.com/SeanNaren/deepspeech.pytorch

10. deepspeech原始项目：Welcome to DeepSpeech’s documentation! - Mozilla DeepSpeech 0.9.3 documentation
11. Windows下安装ubuntu：https://blog.csdn.net/daybreak222/article/details/87968078

做了的事情：

12. 配置环境（windows下的ubuntu、cuda、cudnn、pytorch）
13. 创建的虚拟环境 pytorch_deepspeech2

结果上面的都不好使，应该尝试这个教程：https://blog.csdn.net/weixin_34975139/article/details/112242389

> 可以参考的项目：
>
> 1. https://github.com/yeyupiaoling/PaddlePaddle-DeepSpeech：基于百度开源框架PaddlePaddle实现，配套讲解视频 https://www.bilibili.com/video/BV1Ng411j7xB
> 2. 基于pytorch的：https://github.com/SeanNaren/deepspeech.pytorch
> 3. 官方文档：Welcome to DeepSpeech’s documentation! - Mozilla DeepSpeech 0.9.3 documentation

> 数据集准备：
>
> https://ai.100tal.com/dataset 好未来的开源教育场景数据集
>
> https://wenet-e2e.github.io/WenetSpeech/#download 腾讯+西北工业大学公开的数据集，其中有教育场景

清华镜像：https://pypi.tuna.tsinghua.edu.cn/simple

关于beam search不错的一篇回答：https://www.zhihu.com/question/54356960

一条可能的思路：不重新写ctcdecoder，直接去用warpctc说不定就可以（看一下源码）