**语音识别定时自动测试文档**

1. **概述**

本文档对**自学习算法中定时自动化测试部分**进行详细解释，实现功能：

a) 模型替换：将训练好的声学模型替换引擎的初始声学模型；

b) 生成测试数据：读取数据库音频存储相对路径，拷贝音频到引擎目录下；

c) 测试：对选取的音频采用restful方法测试；

d) 解析测试结果：分别通过restful请求获取识别结果和解析日志获取识别结果；

e) 结果展示：测试结果入库及生成测试报告；

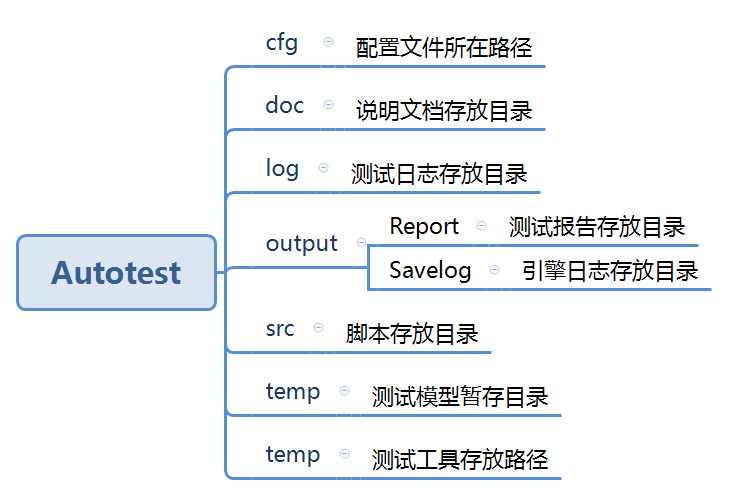
f) 重置：将声学模型重置为初始声学模型以及清除本次测试引擎生成的pcm文件；

g) 定时：定时检测是否生成新模型并对新模型进行测试；

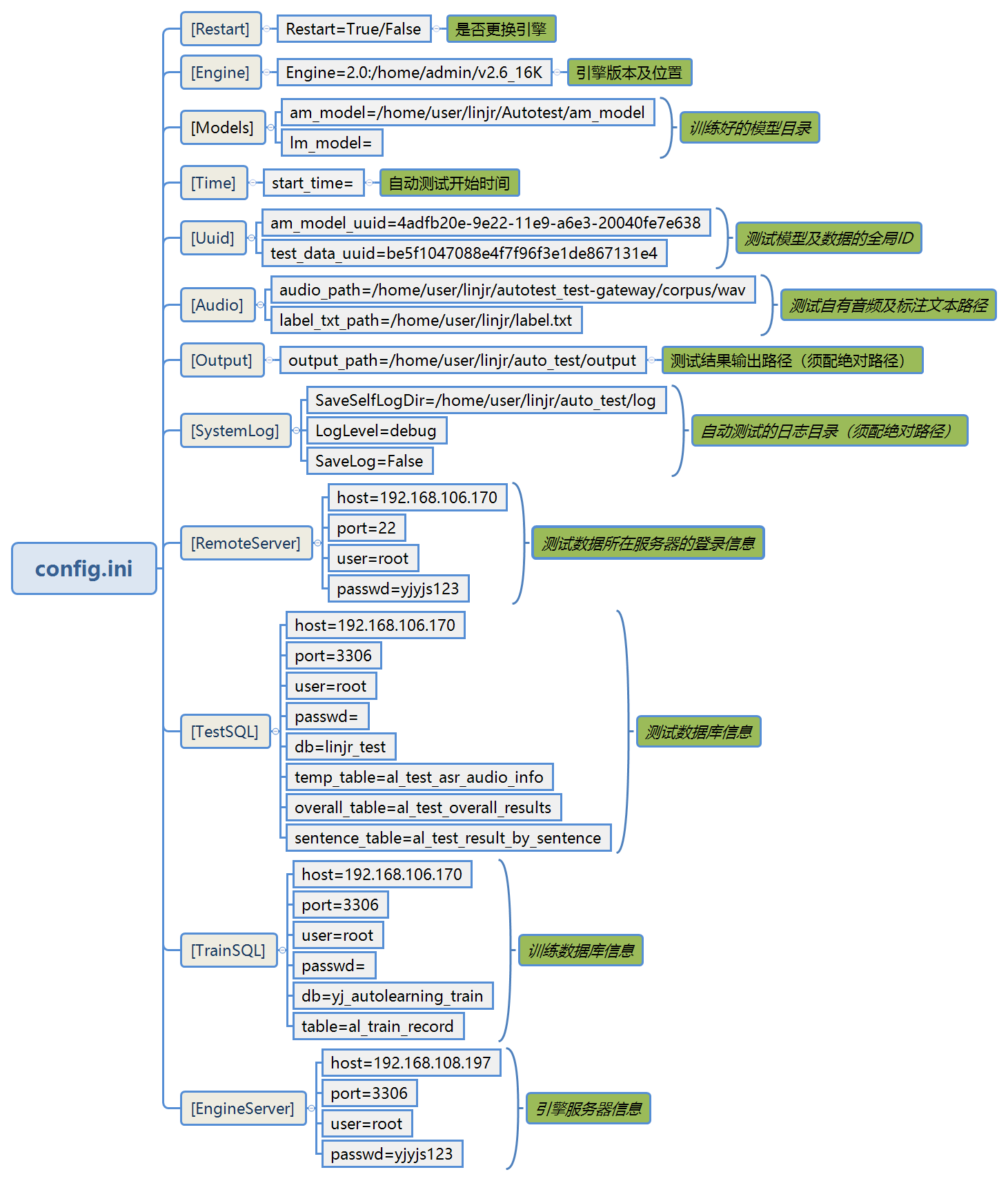
h) 追溯：可追溯每次测试使用的模型和数据；

1. **项目结构**

项目目录位于http://192.168.100.123/gogs/ASR\_BASE/AutoTest.git的ljr分支上,以下为该结构的详细信息解释：



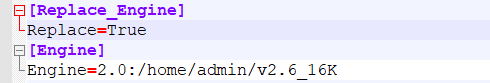
1. **配置参数说明**

****

1. **使用说明**

配置../cfg/config.ini文件

**配置引擎信息**

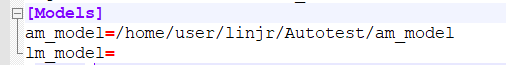


Replace=True:更换引擎； Replace=False:不更换引擎

2.6引擎：Engine=2.0:/home/admin/v2.6\_16K

冒号前表示引擎版本，冒号后表示引擎路径。

**配置模型信息**



am\_model=/home/user/linjr/Autotest/am\_model表示需要替换的声音模型；

lm\_model= 表示需要替换的语言模型；空（不填）表示不替换模型，采用基础模型。

**配置RemoteServer信息**



RemoteServer：数据管理数据库所在服务器的登录信息

**配置TestSQL和TrainSQL信息**

TestSQL：测试数据库信息，其中：

db=yj\_autolearning\_test：测试信息所存数据库

temp\_table=al\_test\_asr\_audio\_info：测试数据音频信息表

overall\_table=al\_test\_overall\_results：存放全部数据的测试结果信息

sentence\_table=al\_test\_result\_by\_sentence：存放逐句的测试结果信息

TrainSQL：训练数据库信息，其中：

db=yj\_autolearning\_train：训练信息所存数据库

table=al\_train\_record：训练结果信息表

**配置EngineSever信息**

EngineServer：引擎所在服务器的登录信息

**执行自动化测试程序**

1）在./src目录下，执行auto\_test.py，会检测TrainSQL中是否有新模型生成，如果没有新模型生成，会在./log/Autotest.log中显示“no new model has been generated!”如果有新模型生成，则开始自动化测试

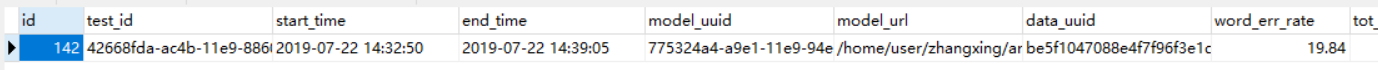
2）自动化测试首先需要等待10-20分钟，进行引擎重启

3）引擎重启成功后，会直接调用prepare\_test\_data.py程序，并从RemoteServer数据管理数据库中获取音频的路径信息，将音频拷贝到引擎中

4）测试数据准备完毕后，会直接进行测试，并将测试结果录入数据库，并在测试结束后，将后续所需的引擎日志拷贝到./output目录，用作日志解析

5）引擎日志拷贝结束后，会先调用parse\_log\_to\_mysql.py程序来解析日志，然后调用calculate\_to\_mysql.py程序，计算测试结果并录入yj\_autolearning\_test数据库的al\_test\_overall\_results和al\_test\_result\_by\_sentence数据表中。其中：

al\_test\_overall\_results存放总测试结果信息



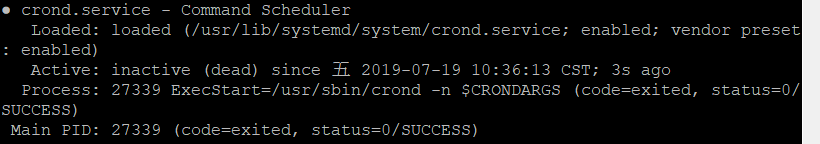
al\_test\_result\_by\_sentence存放逐句的测试结果信息



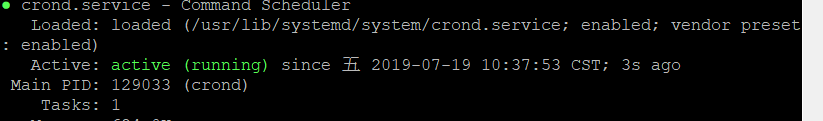
1. 日志解析和测试结果解析完成后，将调用report.py程序，生成测试结果报告，位于配置的output/Report/目录下
2. 最后会调用recover.py程序，将声学模型重置为初始声学模型以及清除本次测试引擎生成的pcm文件
3. **使用建议及说明**
4. 执行./src/auto\_test.py会自动检测训练数据库是否有新模型生成，不需要进行配置模型信息。
5. 执行自动化测试程序中的2）-7）步骤均会自动执行，无需手动执行。
6. 如果需要定时检测训练数据库中是否有新模型生成并对新模型进行测试，可设置定时任务：sudo crontab -e添加定时任务，如下所示：



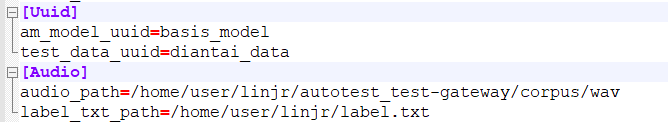
添加任务后可执行service crond status查看crontab服务状态，若服务停止：



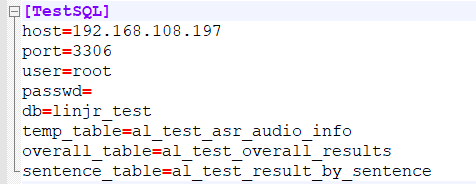
则需重新启动服务：service crond start:



1. 可用al\_test\_overall\_results表里的“model\_url”字段追溯那次测试所用的模型，用“data\_uuid”字段追溯那次测试所用的数据。
2. **自有音频测试使用说明**

测试自有音频时需先配置../cfg/config.ini文件文件中的Uuid和Audio字段，其中Uuid字段给本次测试备注测试模型名称和测试数据名称，Audio字段为自有音频文件夹路径及标注文本路径（标注文本编码与内容格式务必标准），若没有标注文本，可以自行生成伪标注文本，又或只需要对音频进行测试获取引擎日志及识别结果，可以配置“label\_txt\_path=”

对自有音频进行测试时还需配置**TestSQL**为自己数据库，也可配置为作者数据库：



配置完以上信息后执行../src/run\_test.py即可：

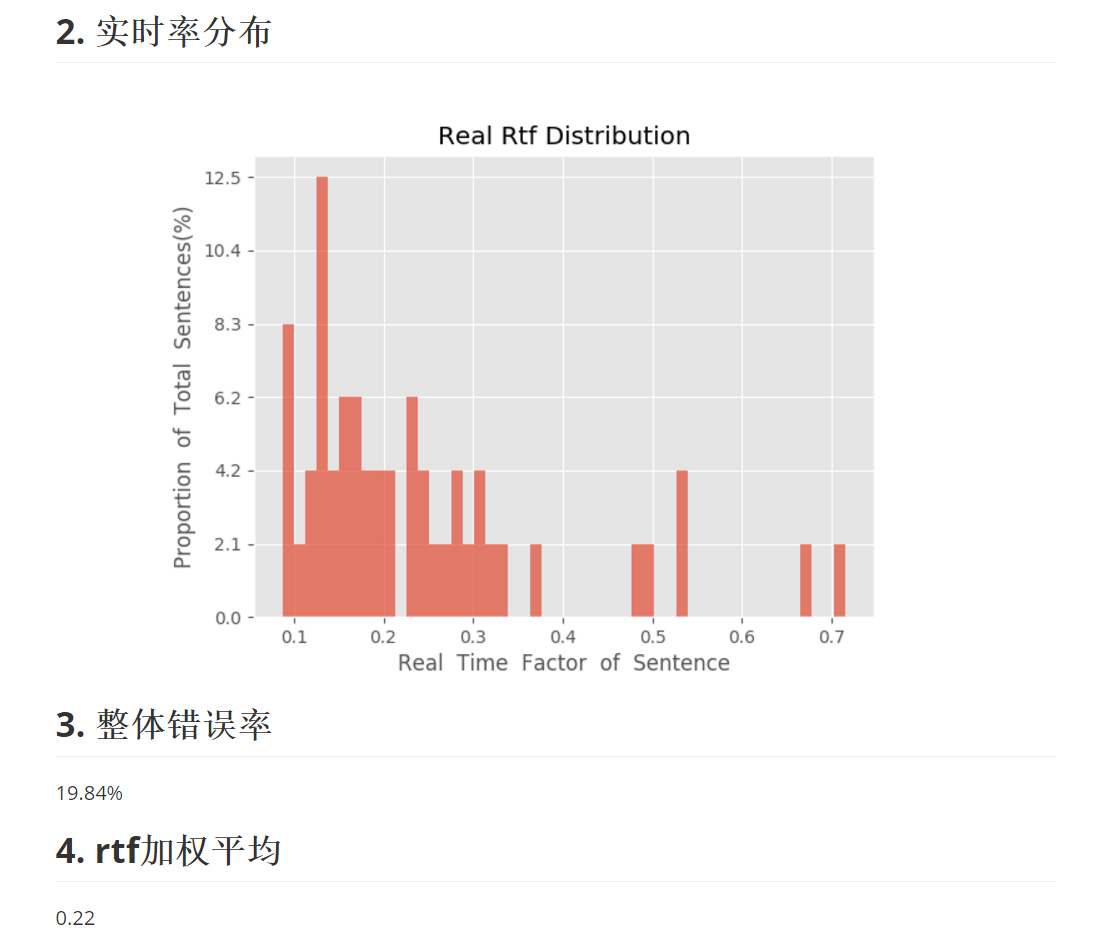
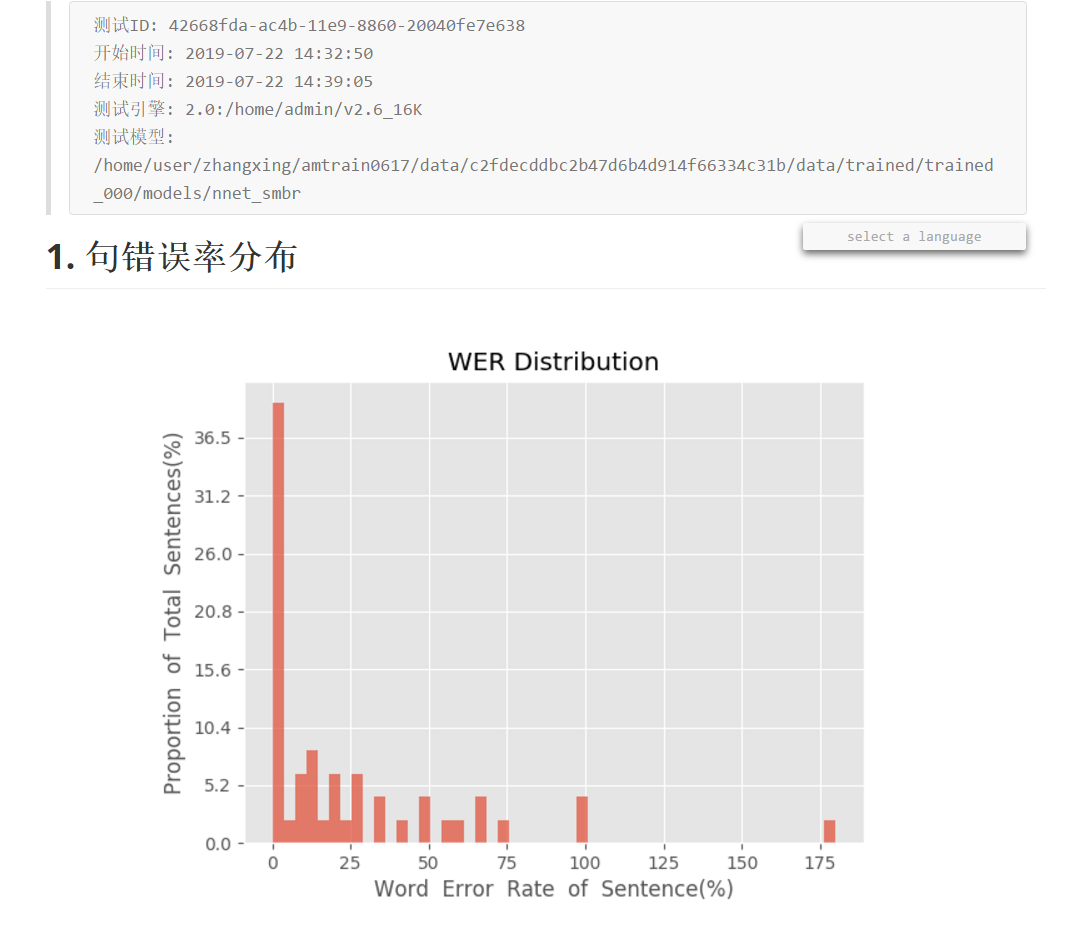
1afcd6f953c0985cb9f4d1d90cad2a8

上图的测试进度表示测试正在进行中。

1. **结果展示**

测试结果分别会在数据库、测试报告以及网页端上展示，数据库展示位于170服务器（自有音频测试位于自配的数据库）yj\_autolearning\_test数据库al\_test\_overall\_results和al\_test\_result\_by\_sentence表，其中al\_test\_overall\_results为总测试结果表，../doc目录下有详细的表结构信息。

测试报告生成在配置的output/Report目录下，测试报告展示如下：



1. **目录src代码说明**

|  |  |
| --- | --- |
| **名称** | **功能** |
| auto\_test.py | 自学习启动脚本 |
| restart\_engine.sh | 重启引擎脚本 |
| abt\_engine.sh | 函数工具 |
| replace\_model.py | 模型替换 |
| run\_test.py | 自测试启动脚本 |
| move\_log.py | 日志转移脚本 |
| read\_global\_config.py | 读取配置信息 |
| prepare\_test\_data.py | 准备数据 |
| parse\_log\_to\_mysql.py | 解析日志入库 |
| calculate.py | 信息计算脚本 |
| calculate\_to\_mysql.py | 计算信息入库 |
| generate\_log.py | 测试日志脚本 |
| report.py | 生成报告 |
| restful\_class.py | restful测试脚本 |
| recover.py | 引擎重置 |
| own\_data\_prepare | 自测试数据准备 |