Record

2020-10-10 同花顺入职备忘录

**Zhaohui Yin**

**Tus, Nov 3, 2020**

<http://172.19.81.34:8090/pages/viewpage.action?pageId=4587828>

1. 工作基础软件

1.1 vanish

* 外网IP：10.10.52.18，内网IP：10.0.68.14
* 账号：yinzhaohui@myhexin.com
* 密码：123456
* 作用：公司内部工作交流软件,https://vanish.myhexin.com

1.2 工作邮箱

* 账号：yinzhaohui@myhexin.con
* 密码：Yzh723787200.

1.3 工作环境

* 主要工具：1） confluence [技术分享]；2） JIRA [项目跟踪]；3）~~gitlab~~ [http://gitlab.myhexin.com]；4）vpn翻墙设置 [链接失效，无法打开]
* 编程相关配置：1）~~内网SVN~~ [并没有添加到内部管理平台]；2）内网服务器创建账号
* 双屏切换：Crtl+Alt+Enter 即可
* 初始密码设置为123456
* 备注：标蓝的区域是暂时无法使用的工具，跟语音入门实验的关系不大，可以暂时先滞后。

2. 常用工具

2.1 内网，外网传输文件

* 外网->内网：\\172.20.54.216\home\视觉与前段\语音项目
* 内网->外网：[暂时不用]

2.2 服务器的使用 \*

1. 需要补充学习的知识：linux命令 [复习] 与shell脚本 [需补充]

2. 连接服务器，开始入门程序测试，主要机器的IP：

1）ssh命令连接服务器: xshell，用户名：yinzhaohui，密码：yzh123456 [内网里面登陆]

2）scp 远程拷文件；3）环境变量设置：/file2/Work/个人目录/.bashrc中填入环境变量设置content ⇒ 在每一台机器个人的bashrc中，source工作目录下的.bashrc文件。

Table1: 常用服务器的IP地址及其功效

|  |  |  |
| --- | --- | --- |
| 常用机器 | IP地址 | 功能与作用 |
| file2 | 10.0.68.102 | 1）/file2/Tools目录下存放公共的工具  2）/file2/Libs 存放了一些库  3）/file2/Work 存放了其他同事的工作 |
| cpu4 | 10.0.68.112 | 声学模型训练、声纹和语种的Work目录 |
| cpu5 | 10.0.68.113 | 语音模型、合成和翻译的Work目录 |
| cpu6 | 10.0.68.114 | 测试机器 |

3. 声纹识别研究 \*

3.1 基础知识点整理

|  |  |
| --- | --- |
| 针对问题 | 基于数据驱动的模式识别问题  **1. 说话人确认技术 SI**（eg. 汽车声控锁）  **2. 说话人辨认技术 SV**（eg. 刑侦） |
| 理论机制 | 1. 数据可行性：本质上是因为声音频率（声腔尺寸 & 发声器官被操纵的方式）各不相同 --> 同一个人同一个字也可能是不同的 --> 不变的是什么？[**模式识别的核心**]  2. 方法可行性：短时平稳，即在一个20-50ms的范围内，语音近似为良好的周期信号。（**周期信号表示可识别**） |
| 不理解之处 | 1. 音调、响度和音色的差异如何表示？（图像的IHS都可以计算并表示）   * 音调与频率相关，频率越高音调越高。 * 响度与振幅相关，振幅越大响度越大。 * 音色则是为不同频段能量的差异，音乐和噪声差别很大   2. 20-50ms长度的语音如何映射为39-60维的向量？   * 本质上是对于语谱图\*的理解。 |
| 识别精度受控因素 | 1. 背景噪声（音乐、风声，收集器的差异，**多人说话**等）  2. 个人原因（身体情况、**情绪**等） |
| 痛点 | 1. 与发音者从根本上相关的特征是什么，如何提取？ （**特征提取**）  ---> 具体表现：文本无关的语音识别；  2. 复杂背景下的语音识别？（**信噪比小**）  ---> 具体表现：远场景语音识别[信号弱]，嘈杂环境语音识别[噪声大]，多人说话。 [多信号源，互为信号or噪声] |

语谱图（Sonogram）通俗来说是对语音信号连续地进行频谱分析得到的二维图谱，横坐标表示时间，纵坐标表示频率，每个像素的灰度值大小反映相应时刻和相应频率的信号能量密度。每一时刻都可以用这个时刻附近的一短段（15ms）语音信号分析得到一种频谱（因为短时平稳，周期性信号，有较为稳定的频率）

|  |  |  |
| --- | --- | --- |
|  |  |  |
| （1）A：“四” | （2）B：“四” | （3）A：“九” |

3.2 研究进展

Table2 目前声纹识别领域的研究进展

|  |  |  |  |
| --- | --- | --- | --- |
| 时间轴 | 方法 | 方法原理 | 创新点及不足 |
|  | 模版匹配 | 信号比对，要求内容相同 | **与文本有关** |
| 2000 | 基于高斯混合模型的声纹识别算法（GMM）\* | 统计学习理论 | **与文本无关**，但是所需时间过长 |
|  | 联合因子分析法（JFA） | 分别建模说话人空间信道以及残差噪声，每一步引入误差 |  |
| 2010 | iVector/PLDA\* | 把语音映射到一个固定且低维度的向量上 | 意味着所有机器学习的算法都可以用来解决声纹识别的问题 |
|  |  |  |  |
| 2011 | DNN-based speech\* recognition | DNN可以从大量样本中学习到高度抽象（意味着无法解释）的音素特征 | 识别率提高了30%，对噪声有很强的免疫力 |

目前的困境：动态变化的发声器官与声音，稳定性不如人脸及图像。其重点就在于是否能够轻易的被相似特征的声音攻击。根本问题还是在于：我们耳朵区分不同人的声音到底是因为什么机制？[属于生物识别技术]

3.3 开展入门实验 \*

利用python生成语谱图，分别录入“四”、“十”、“五”的音频，以及对比“四十五”和“五十四”的区别。结果发现确实是存在规律，无论是机器翻译还是语音识别的理论根基都是语谱图。

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | |  |
| “四” | “十” | | “五” |
|  | |  | |
| “四十五” | | “五十四” | |

在此基础上，综合考虑人耳的听觉属性（低频敏感，高频稍弱）、噪声鲁棒性、均衡不同频段的能量差异设计合适的声学特征（信号处理层面的变换）。传统的声学特征包括梅尔倒谱系数MFCC、感知线性预测系数PLP、深度特征Deep Feature、以及15年公开发表的能量规整谱系数PNCC等，都能作为声纹识别在特征提取层面可选且表现良好的声学特征，一般这些特征可以被映射为39-60维的向量。（图像中HOG也有128维）。

## 入门实验的设计思路【技术性内容】

1. 直接在自己电脑上进行程序设计；**git->tools/->src/**

\* kaldi 下载及编译：https://zhuanlan.zhihu.com/p/44483840，速度巨慢,安装时务必小心

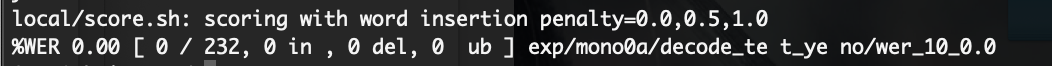
\* BUG解决方案：https://zhuanlan.zhihu.com/p/141260586，**检查依赖时会出现问题**

https://blog.csdn.net/mimiduck/article/details/107383684，**cub下载问题**

https://www.jianshu.com/p/6f650badc462，**homebrew update问题**

2. **利用yesno数据集来进行识别，kaldi**

\* 实验结果：结果表示没有错误



**\*** 代码解读：理论 & 代码

链接：https://blog.csdn.net/cg\_speech/article/details/81202694?utm\_medium=distribute.pc\_

relevant.none-task-blog-title-2&spm=1001.2101.3001.4242

3. **利用标准数据库TIMIT进行声纹识别的实验**

\* 数据集内容如下表所示：

Table3 TIMIT数据集介绍

|  |  |
| --- | --- |
| database | description |
| TIMIT | 声学-音素连续语音语料库，采样频率16kHz |
| 包含内容 | 6300句子（630\*10，2方言/5紧凑/3发散），音素级别进行分割。其中70%是男性 |

链接：https://www.zhihu.com/search?type=content&q=%E5%A3%B0%E7%BA%B9%E8%AF

%86%E5%88%ABkaldi。

⚠️ 注意事项：

2. 连接公司的服务器，进行程序测试

\* Xshell 连接服务器

3. 跑出结果

Figure1 程序测试结果

4. 语音识别

声纹识别是识别说话人的身份，语音识别是识别语音的内容。语音识别从不同人的词语中寻找共同的因素。技术较为成熟，重点是声音的录入、及内容的识别准确性。

|  |  |
| --- | --- |
|  |  |
| (a) 声音录入 | (b) 内容识别 |

附录一:基础知识补充学习ssh

1.1 Linux命令以及 shell脚本学习 [课后学习即可]

1. 目的：连接服务器以及在服务器上进行相关的操作。

2. 链接：http://billie66.github.io/TLCL/book/

3. 重点学习内容：

|  |  |
| --- | --- |
|  |  |
| (a) | (b) |
|  |  |
| (c) | (d) |

Figure1: linux命令与脚本学习网上学习资料指南，其目的是为了更好的进行服务器上面的操作。

1.2 CUDA编程（？）

附录二. 部门研究方向及成员

|  |  |  |  |
| --- | --- | --- | --- |
| 研究方向 | 成员 | 应用 | 分类 |
| 前端 | 叶帅帅、王培尧 | DiHard实验、语音增强、麦克风阵列算法 | 前端 |
| 机器翻译 | 王海波、李高鹏、秦波、斯拉吉丁、 | 1. PC+手炒--美股公告翻译  2. PC+手炒--F10翻译（？）  3. PC -- iFinD英文版  4. PC -- 翻译渗透  5. 手炒 -- 直播服务语音翻译 | 机器翻译 |
| 语音识别声学模型训练 | 陈顺飞、项秉伟、石梦、陈龙、张琪 | 1. 数据标注，数据库建立  2. 语音增强  3. 与前端联合，识别系统  4. 四川话、东北话识别[方言识别]  5. 模型优化【实时性 or 性能】 | 语音识别 |
| 语音识别语言模型训练 | 杨磊、谷宾宾、陈龙、吴浩、 | 1. 文本数据扩充  2. 基于神经网络的语言模型研究和应用  3. 文本生成模块  4. 识别文本纠错模块  5. 标点生成模块  6. 语言模型的更新  7. 更多开源框架工具的应用转化 |
| 语音合成 | 鲁泽如、陆建、张鹏 | 产品层面提供更多声库选择 | 语音合成 |
| 声纹识别 | 喻陈毅、王金龙 | 1. 训练8k中文xvector声纹模型，用于电话场景下的声纹识别  2. 测试并优化16k的xvector声纹模型，根据需要不断优化  3. xvector声纹引擎测试与优化  4. VAD判断  5. 说话人确认SVC2019 | 引擎系统 |
| 引擎（知识盲区） | 陶驰雷、王贵山、刘小康、陈禹行 | 1. 语音合成（陶）  2. |
| 语音识别后处理（？） | 王鼎 | 语种识别 |
| 情感识别 | 李兴风 | 1）新方向扩展，展开相应的学术研究  2）对话机器人 | 新方向 |