

# 让掌控板也能识别语音

## ——对听障助手项目的进一步改造

谢作如 浙江省温州中学  
程龙恺 上海人工智能实验室

**摘要:** 本文对“AI助听器”项目进行深度分析,挖掘了掌控板内置麦克风的功能,设计了利用掌控板录音、行空板推理的工作方式,并给出了标准化的参考代码,以为中小学AI科创项目的开发提供新方案。

**关键词:** MMEdU; 深度学习; 语音识别; AIoT

**中图分类号:** G434 **文献标识码:** A **论文编号:** 1674-2117 (2023) 09-0073-03

笔者自从开源了用时频谱图(语谱图)来识别语音的项目后,激发了好多创客们的创造力,类似的离线版本的语音识别项目不断涌现。这类项目的核心工作方式是先将音频转换为时频谱图,再用图像分类的方式来识别简单的语音。其优势在于不需要借助类似百度、科大讯飞的语音识别接口,离线状态下也能工作。而面向听障人群的“AI助听器”项目选择了掌控板和行空板这两款常见的开源硬件来实现语音识别和信息提示,获得了创客们的好评。

### ● 原项目的不足和改造方案分析

在“AI助听器”项目中将行空板作为模型推理服务器,掌控板作为信息显示终端。先通过各种声音训练出深度学习的模型,再部署到行空板上,然后利用行空板的麦克

风采集声音,通过物联网将相应文字信息发送给终端。通过文字显示、灯光加震动提醒,让听障人“看见”声音或者感触到声音。这个项目虽然很酷,但经过分析后笔者认为还存在几点不足:

①模型的推理速度较慢。项目使用了Keras的H5模型,这个模型使用在线平台训练,网络结构是MobileNet,在行空板上推理的速度有点慢,约为0.9秒。最让人难以承受的是,用Keras载入并初始化这一模型需要很长时间,约为1分钟。经过测试,如果采用ONNX模型完成一次推理约为0.5秒;NCNN模型则会更快一些,在0.2秒左右。而ONNX和NCNN模型的初始化是很快,1秒钟内都能完成。

②声音采集终端选择不合理。项目中使用行空板来采集声音,识别的范围只能是行空板的周围。一般

而言,采集声音的终端应该是多个的,如需要关注的某个房间、厨房、大门口,以及“听障人”自身所在的区域等。因此,采集声音的终端应该具备价格便宜、方便移动的特点。

③服务器端的代码有些复杂。因为服务器端要完成多种工作,既要采集音频,又要转换数据,还要对H5模型做推理,发送MQTT信息。代码一长,调试代码就特别复杂,这个项目的可推广价值也会受到影响。

笔者对原有方案进行了改造,核心工作是用行空板作为模型推理服务器,掌控板作为声音采集和信息显示的终端,其中AI模型采用ONNX。这个AI应用系统采用了1块行空板+n块掌控板的方案,用MQTT连接起来,如下页图1所示。多块掌控板和行空板协同工作,形成一个完整的物联网语音读取、传

输、识别和反馈的设计流程。

### ● 掌控板端代码的编写

掌控板作为终端(也可以称为下位机),主要任务是实时检测音频,并通过MQTT协议发送至行空板,实时监视服务器传送的消息,根据不同的指令执行相应的工作。笔者设计返回的消息是一个列表,如["'sms','敲门'"]表示有人敲门,["'sms','哭声'"]表示有孩子在哭。然后根据不同的消息内容执行不同的工作。

让掌控板用MQTT的方式传输音频是一道难题。虽然一开始笔者就想到用Base64的方式编码的方法,掌控板编码,服务器端再解码,但没想到长一点的音频用MQTT方式传输系统容易崩溃。为了解决这一问题,笔者一度采用了EasyAPI,即在行空板上搭建Web服务器,用HTTP方式来传输音频。可喜的是,经过多次测试后终于找到了分批传输的形式,解决了这一难题。其原理并不复杂,是将一个数据包分成多个较小的数据块进行传输,同时利用校验和序列号等机制保证数据的完整性和正确性,以提高数据传输的可靠性。

笔者借助了Python内置模块binascii来实现。binascii是二进制数据和ASCII字符串之间转换的函数,用于将音频字节码转换为base64编码的ASCII字符串。核心参考代码如图2所示。

这段代码的作用是将大型

的音频文件分成若干个较小的块,每个块大小为step\_size,然后通过MQTT协议发布到指定的主题“IOT\_pubTopic”。在每次循环迭代时,使用binascii.b2a\_base64()函数将当前块转换为base64编码的ASCII字符串,其中min(index+step\_size,len(audio\_bytes))用于确保最后一个块小于等于step\_size大小,以避免在分块时发生错误。

考虑到行空板每进行一次推理都要耗费0.5秒的时间,而终端(掌控板)提交的音频可能存在“空白”的音频,笔者还设计了音频实时检测的功能。当检测到声音高

于设定的阈值时(如图3),才开始记录音频数据,并在录制结束后将音频发送到服务器进行推理。

### ● 行空板端代码的编写

在这个系统中,行空板作为服务器(也称上位机)实时监听MQTT消息,将有效的语音转为图像后进行推理,再发送识别后的结果或者相应的命令。行空板除了能执行模型推理的工作外,还内置了SIoT服务器软件。SIoT是一款使用非常方便的MQTT服务器,不用设置就能直接使用。

笔者采用MMEdU训练人工智能模型。将音频批量转换为图片后用MMClassification模块训练,

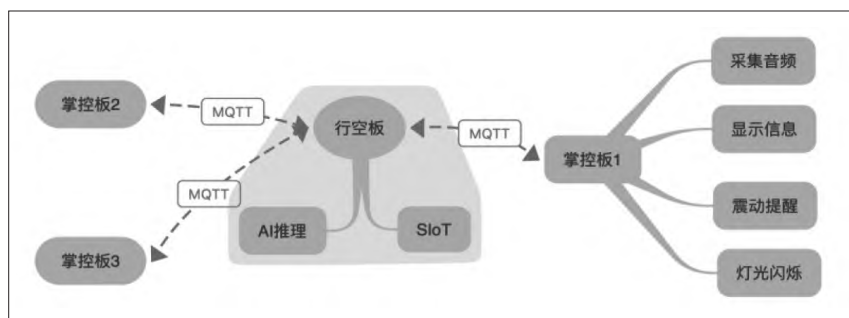


图1

```

import binascii
for index in range(0, len(audio_bytes), step_size):
    b =
    binascii.b2a_base64(audio_bytes[index:min(index+step_size, len(audio_bytes))])
    siot.publish(IOT_pubTopic, b)

```

图2

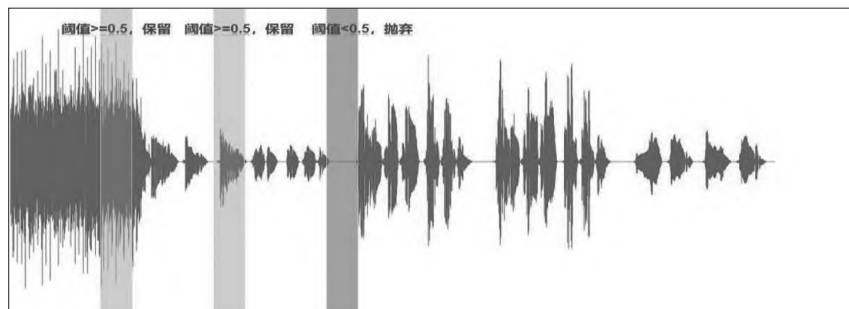


图3

```
def EasyAIoT(client, userdata, msg):
    global message_tol
    topic = msg.topic
    payload = msg.payload.decode()
    receive_msg = binascii.a2b_base64(payload)
    if isinstance(receive_msg, bytes) and len(receive_msg) > 0:
        message_tol += receive_msg
        if len(receive_msg) != step_size:
            tmp_message = message_tol
            message_tol = b''
            # 将接收到的结果保存至本地并推理
            file_save(tmp_message)
    else:
        message_tol = b''
```

图4

```
from BaseDT.data import ImageData
new_img = ImageData(image, backbone = 'MobileNet')
```

图5

也选择MobileNet网络模型,然后使用model对象的convert方法来转换为ONNX模型。这一过程在本栏目之前的文章中已经介绍,不再赘述。核心工作是编写了一个名叫“EasyAIoT”的函数,实时检测指定的MQTT消息,并将分批的消息内容保存为音频文件再推理,采用了siot库(一个简单的MQTT库)和onnx推理框架(如图4)。

为了让数据预处理变得更加简单,笔者使用了BaseDT库。BaseDT库是专为人工智能的数据

预处理开发,其中ImageData内置了常见的计算机视觉模型的转换参数,输入SOTA模型的名称(如MobileNet)就能完成预处理,代码非常简洁。参考代码如图5所示。

### ● 新项目的测试反馈和总结

在完成项目后,笔者对这套新的“AI听障助手”系统进行了实地测试。整体效果还不错,行空板能够同时连接多块掌控板,运行速度稳定,甚至可以借助其内置的热点功能,直接和掌控板形成了一个完整的、闭环的智能物联网系统,即

AIoT。虽然掌控板不能直接部署AI推理模型,但是用内置的麦阵设备录制音频,再借助其他开源硬件进行识别的方式和常见的智能音箱是一致的。

笔者再次分析了原来的听障助手项目,认为作者之所以没有采用掌控板来采集声音,很可能是误以为掌控板内置的采集音频设备是普通声音传感器。实际上那是一个微型的麦克风,录制音频的效果虽然不如计算机,但用于简单的语音识别已经绰绰有余。行空板作为服务器,掌控板作为终端,使用物联网技术将二者结合起来,这将是中小学创客教育中最常见的硬件组合方案。为了降低人工智能模型推理和部署的门槛,笔者会将这一项目的核心代码继续优化,最终将实现用图形化的方式来启动能够监听消息并进行实时推理的物联网应用。

### 参考文献:

- [1]谢作如,胡君豪.让行空板变身为能识别方言的智能音箱[J].中国信息技术教育,2023(01):93-95.
- [2]吴俊杰,戴娟,谢作如.中小学AI教育需要怎样的学习工具[J].中国信息技术教育,2022(12):4-10.
- [3]谢作如,程龙恺.当MMEdU遇上行空板——“智能稻草人”项目的后续研究[J].中国信息技术教育,2022(23):77-79.
- [4]云天.行空板——“AI助听器”,让“听障者”看到声音[EB/OL].<https://mp.weixin.qq.com/s/LT0M-d4BmgW36BCF-PahIA>.