

ECGR 4127/5090 项目 2:关键词识别

v1.0

由于 2023 年 5 月 5 日

在本项目中,您将使用 Arduino Nano 33 BLE Sense (或 Pico4ML 板)。您将在 5 月 5 日的期末考试期间展示您的作品。

规格

- 该项目将单独或两人一组进行。在小组之间,您可以讨论概念,但你不应该共享代码。
- 模型应该至少识别两个词,其中至少一个应该是您选择的自定义词(例如您的名字、您最喜欢的颜色、您的宠物的名字)。它不应该是语音命令数据集中的“Google 30”词之一。每个团队必须选择一个独特的词。您的第二个词可以是语音命令数据集中的词之一,也可以是您选择的另一个自定义词。
- 为了训练您的模型,您将收集目标词的几个话语。我们鼓励您合作收集训练数据,例如,您可以记录另一个团队说出您的目标词,然后他们可以记录您说出他们的目标词。您可以使用 embeddd 板本身作为录音设备,也可以使用其他录音硬件,例如电话、笔记本电脑等。
- 部署的模型应该在流场景中工作。换句话说,它应该检测来自麦克风的连续语音流中的目标词,而不是简单地对离散的音频帧进行分类。

表 1:分级

标准满足规	重量备注	
格 35% 您是否满足此处列出的最低规格?		
表现	20% 最大化绩效的方法。	
勘探	10% 除了基线示例中包含的内容之外,您在多大程度上尝试了新方法?	
报告清晰度	20% 解释、清晰易读的数字、语法、格式、ETC。	
第一阶段提交	10% 第一阶段上交的清晰度和完整性。	

报告内容及格式

在报告中记录放大器的设计和测试。您可以选择将报告格式化为 IEEE 期刊样式。图片应包含在文件正文中,而不是单独打印并钉在背面。

你的报告应该彻底解释训练和部署你的模型的过程,以及你的实验结果。以下项目代表报告的关键部分,但这并不构成您的报告应解释内容的详尽清单。报告应包括以下部分:

介绍描述项目的目的,你打算完成什么,并告知读者
即将到来的部分。两到三段就足够了。

模型架构

- 您如何选择神经网络的拓扑结构?网络的规模是否受限于
参数数量、运行时间或两者? · 你选择了哪些输入
特征?你是如何选择它们的? · 您进行了哪些实验来探索其他架构? · 包括任何可能
有助于解释模型设计的图表或解释。

数据和培训 · 您是如何收集数据的?您是否与他人共享资源?你用的是什么硬件?您是否使用了任何其他数据源? · 你采取了什么措施来确保你的训练数据涵盖环境的自然变化,

语音、口音等? · 如果

有的话,你做了什么增强? · 您是否做过任何实验来估计您需

要多少数据? · 您是否采用了任何特定的训练技术,例如蒸馏、迁移学习或合成数据

一代?

- 您认为有助于解释您的数据采集和培训的任何其他图表或信息
管道。

结果

- 在软件中报告模型的准确性,即在TensorFlow 中对训练、验证和测试集进行测量。

- 描述模型在其部署的流上下文中的性能。它是否响应目标词。它是否对其他声音表现出误报?是否有任何特定的声音或单词比其他声音或单词
更能引起误报 (例如近同音字)。它是否在任何特定的说话者群体或任何特定的演讲风格上更频繁地失败 (例如,当说话缓慢或对于一种性
别的说话者等时,模型可能会失败)

- 推理需要多长时间才能运行?相关地,您对部署的模型使用什么采样率 (每秒帧数)? · 以几种不同的音量播放5-10 分钟的噪声源录音,
并记录误报的次数。转化为每小

时的误报率。使用电话和分贝计应用程序测量麦克风的声级。

汇总表将以下主要结果合并到汇总表中。

- TensorFlow 中训练、验证和测试数据的准确性。 · TensorFlow 中语音命令数
据集中单词的错误拒绝率 (FRR)。 · TensorFlow 中的 FRR 用于您的自定义词。 · 参数数量和MAC。 · 输入张量形
状。 · 采样率 (又名帧率)。 · 误报率 (FA/小时)。

讨论总的来说,您将如何评估已部署模型的性能?它看起来像一个可用的语音控制系统吗?哪些改进可以帮助它成为一个更有用的语音界面? (例如,“它
需要对安静的声音做出更灵敏的反应。”或“它需要更少的错误警报来响应电视。”)如果您的模型或结果中有任何奇怪的行为,请描述它们并尝
试解释它们。遇到了哪些困难?你会做些什么不同的事情?

结论简要总结你的结果,你学到了什么?

参考资料如果您使用的信息来自课程之外的任何论文、应用笔记、网站或教科书
文本,在这里引用那些作品。