

揭开下一代人工智能和嵌入式系统的面纱

作者 Stuart Cording(Elektor)

译者 君谦

0 引言

人工智能(AI)通常是一件通用的解决问题的利器,随时准备迎接任何挑战。而有些时候,它被认为预示着文明本身的终结。作为工程师,我们如何才能能在开发应用程序的过程中更好地利用人工智能,或者更正确地利用机器学习(ML)呢?人工智能领域的最新进展真的会让我们变得多余吗?

提到人工智能(AI),无疑是吸引主流媒体注意的一种重要方式。由于互联网和云服务的发展,再加上时不时有问题的数据源,新的人工智能驱动的服务似乎无处不在。点击几下鼠标后,你的文字可以放进 Morgan • Freeman 的嘴里,或者你的图像融入到一件新的艺术品中,这些有趣而发人深省的想法,要在嵌入式系统中实现并非易事,但这是该行业内部正在发生的事情。

大多数嵌入式系统使用基于规则的编程方法来实现其功能。给定一组输入数据,一系列 if/else 或 switch 语句决定如何响应。对于有限数量的输入,这种方法效果很好。然而,在某些时候,输入的数量或它们之间关系的微妙之处使得很难定义明确的、程序化的规则。

举个例子,想象一下工厂里有一台马达一周 7 天、每天 24 小时运转。经验告诉我们,随着时间的推移,会发生磨损,轴承润滑脂会变稠,进而改变机器的启动时间、噪声、振动模式、电机工作温度和电流等参数。定期维护是当今应对这一挑战的方式,但它会导致固定的停机时间,从而中断生产。

此外,它不太可能捕捉到由轴、轴承、套管或夹具的细微断裂引起的即将发生的故障。经过正确训练后,人工智能方法可以使用复杂的数据源组合来确定潜在故障。如果这种智能可以部署到基于微控制器的系统中,那么就会得到一个经济实惠的监控系统,既节省了时间和金钱,又减少了不必要的润滑和零件更换造成的浪费。

1 将智能引入微控制器中

在微控制器层面,我们谈论的是机器学习(ML),而不是 AI。这意味着对机器进行编程,使用通过分析可用数据制定的规则来做出决策。虽然微控制器的功能强大到足以执行这样的 ML 算法,但从训练数据中学习仍然超出了它们的能力范围,如果不是云服务器,至少需要一台台式计算机来执行。Edge Impulse 成立于 2019 年,开发了

一个专门用于嵌入式系统中 ML 的平台,并成功地与全球半导体供应商进行合作推广。

任何 ML 应用程序的起点都是数据。虽然一些应用(比如自动驾驶)需要 TB 级的训练数据,但简单的基于微控制器的系统可以从少量(几千字节)的数据中学习。因此,将数据从板上取出并放到 Edge Impulse 环境中是第一个挑战。最初的想法是使用 Arduino 的串行接口将其发送到您的 PC,然后从那里将其作为文本文件上传。然而,他们的平台是为了直接获取数据。

2 数据:AI 的食物

其中一个工具是 Data Forwarder,这是一个命令行(CLI)应用程序,可以将数据从开发板直接发送到 Edge Impulse 环境。使用用户名和密码即可在 PC 上的串行端口和服务器之间建立链接。在微控制器方面,所需要的只是通过串行接口以逗号或制表符分隔的格式输出数据。如果采样率相对较低,那么这是直接从传感器收集代表性数据的理想方法(见图 1)。更强大的嵌入式系统(如树莓派或 NVIDIA Jetson Nano)可以使用提供的软件开发工具包(SDK)。它也支持麦克风和摄像头等传感器产生更可观的数据量。

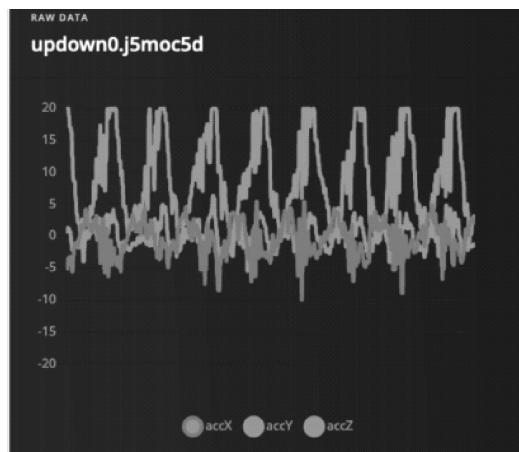


图 1 三轴加速度计提供运动数据,以开发基于 ML 的手势识别应用程序

数据上传后,下一步就是定义“脉冲”。它们由两个部分组成,第一部分将数据切片成更小的块,并使用信号处理技术提取特征,这确保了可用的传感器数据被转换成第二个信号处理阶段信息的一致性;第二部分是学习和分类

的过程(见图 2)。在一个示例项目“连续运动识别”中,很好地解释了如何配置这些块来分析加速度计数据并将此输入分类为 4 种手势之一。

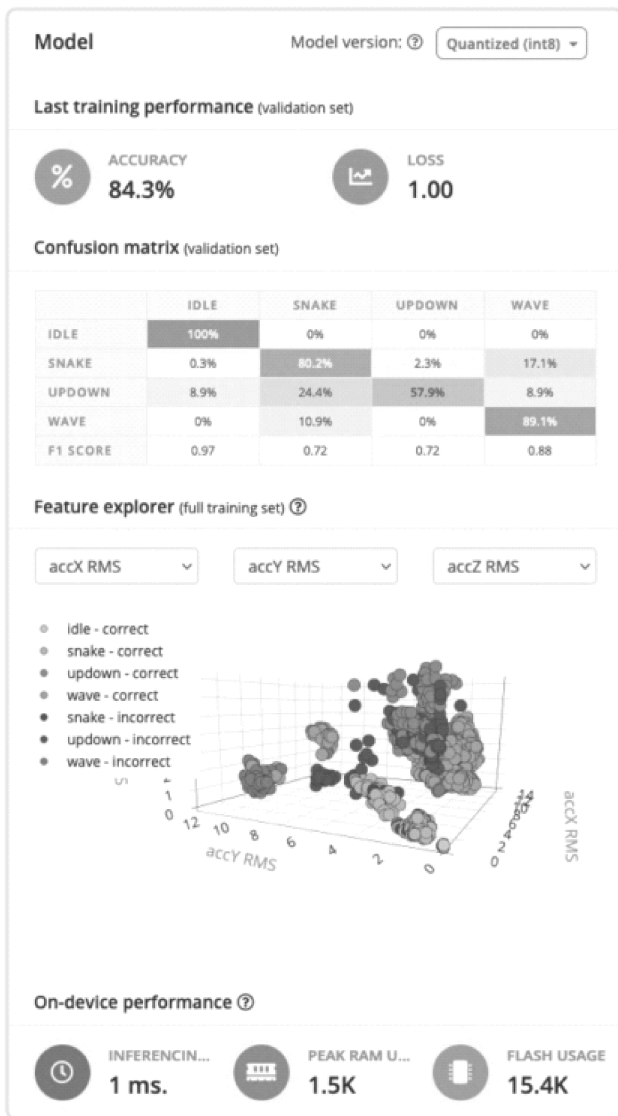


图 2 Edge Impulse 环境在初始训练周期后提供关于准确性、速度和内存使用情况的反馈
(来源:Edge Impulse)

这是任何 ML 开发中最关键的一步,通常需要横向思维和多次迭代来确定最佳方法。有时候,忽略一些传感器输入是最好的解决方案,而其他时候,则需要更多的数据。你甚至可能会发现正在过度学习,或者选择的神经网络模型不适合你试图进行的分类。另一个至关重要的步骤是异常分类。在示例项目中,有 4 种定义的手势。然而,需要排除其他类似于学习手势的动作。良好的异常检测提供了鲁棒性更好的 ML 结果。

最后一步是部署。通过 web 界面下载所选设备的固件,以便集成到你的应用程序中。对于 Arduino,会生成

一个库,而对于其他微控制器,则可以生成一个 C++ 文件。当然,微控制器的性能差别很大。为了确保最佳结果,Edge Impulse 提供了 EON 调谐器。该工具可以通过使用目标设备、内存大小和延迟等信息进一步提高检测精度,加快推理速度,并降低内存需求(见图 3)。

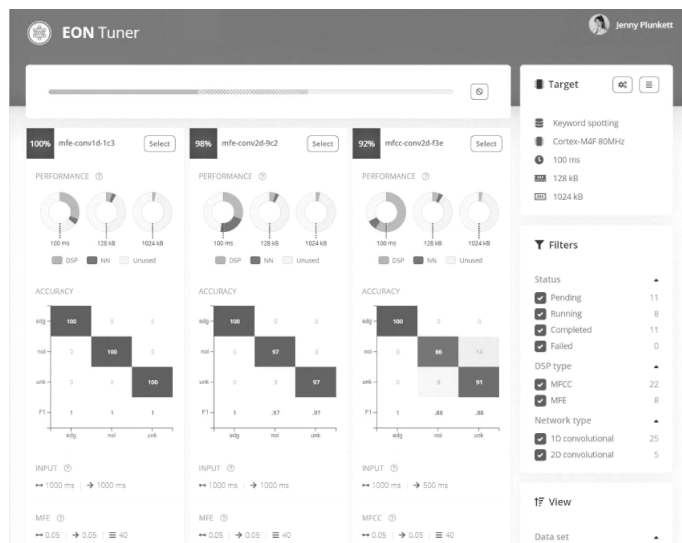


图 3 使用 EON Tuner 对目标微控制器进一步优化了 ML 算法,从而提高了推理响应速度并减少内存使用(来源:Edge Impulse)

3 实际应用中的 ML

实际的应用程序正在使用这种方法来嵌入 ML 功能。SlateSafety 的 BAND 集成了一系列生物识别传感器,以监测在具有挑战性的情况下工作的工人(见图 4)。这些人员包括从急救人员到穿戴重型个人防护装备(PPE)的工业工人,比如消防员。他们的产品通常会将数据上传到云端,这样就可以监测用户的生命体征了。



图 4 SlateSafety 使用 Edge Impulse 提供边缘 ML 生理监测并改进 BAND V2 安全监测(来源:SlateSafety)

但是,特别是在灾难情况下,网络连接可能不完整甚至根本不存在。开发团队利用 Edge Impulse 将边缘 ML 集成到现有产品中,并对历史生物识别数据进行了训练。使用

EON Tuner,该算法针对硬件进行了优化,然后使用无线更新进行部署。现在,即使在没有无线连接的情况下,BAND也可以在存在热衰竭风险时向佩戴者提供警告。

4 利用 AI 改进产品开发

当然,人工智能并不一定要集成到产品中,它也可以用于产品的开发。今天,许多复杂的应用程序都是使用基于模型的方法开发的,本质上是使用软件和物理学的数学方程来描述事物是如何工作的。然而,这种方法也有其局限性,这就使 Monolith 和他们的自学 AI 平台有了用武之地。

该平台能够根据已经收集到的数据学习复杂系统的物理特性。例如,车辆在测试轨道上进行一系列测试,多个传感器监测偏航和滚转以及车轮速度和加速度。收集不同悬架刚度的数据,可以很好地了解车辆对一系列驾驶情况的反应。通常情况下,这些数据将被分析,从而在下次试车时应用新的设置。Monolith 可以评估第一组测试运行的数据,并以很高的精度预测悬架变化的结果。该结果可用于更快地磨练最佳悬架设置,减少所需的额外物理测试运行次数。

这种方法也可以应用于计量。燃气表必须非常准确,以确保正确计费。但当仪表必须测量一系列不同的气体时,这是具有一定挑战性的。对客户来说,模拟超声波仪表已经将纯粹的数学分析推向了极限,使得重复的测试过程成为实现所需认证的唯一校准解决方案。幸运的是,所有的测试都产生了丰富的数据集用于分析。使用自学习 AI 模型,所需的测试量下降了高达 70%,显著加快了开发速度。

5 大幅压缩 AI 计算成本

像 DARPA 大挑战赛这样的比赛,团队建造了可以穿越蜿蜒路线的自动驾驶汽车,引发了人们对自动驾驶汽车的兴趣。近 20 年后的今天,已经投入了大量资金,但收效甚微,似乎只有 Waymo 提供了真正的自动驾驶汽车乘车服务,但这些服务仅在美国的凤凰城和旧金山运营。其中一个问题是,让一台计算机控制汽车是非常具有挑战性的。车辆不仅要不断评估周围的情况,还必须预测其他司机和道路使用者的行为,比如行人和骑自行车的人可能不遵守交通规则。

目前的情况是,汽车的电气和电子(E/E)架构正在发生变化,以满足未来自动驾驶汽车的需求。随着大量传感器提供大量数据,行业正在转向自动网联车。目前,这种方法正在形成先进的驾驶员辅助系统(ADAS),它通过控制制动、加速和转向,可以在驾驶员犯错时介入。根据美国汽车工程师协会(SAE)的车辆自动驾驶水平,高档车辆目前达到了 L2+级,其中一些达到了 L3 级。然而,完全的“解放双手”级别的自动驾驶是 L5 级,所以我们还有很

长一段路要走。

Eurotech 等公司正在支持该行业加快必要算法的开发。目前,一次 8 小时的试驾会收集 120 TB 的数据,这些数据必须返回实验室进行分析和处理。AI 算法的改进可以在实验室中使用收集到的数据进行测试,但支持现场测试和算法开发的数据很少。

利用在液体冷却方面的经验,Eurotech 提供了一系列能够胜任这项任务的边缘 AI 硬件,本质上是放在汽车后备箱里的小型超级计算机。像 DYNACOR40-36 这样的设备经过加固,可以在公路和越野车辆中使用。这款无风扇计算机采用 16 核英特尔至强 CPU(64 GB RAM),最多两个 NVIDIA GV100 GPU(32 GB RAM),提供 237 TFLOPS 来处理深度学习应用程序(见图 5)。几个千兆以太网接口支持摄取大量传感器数据,包括雷达、摄像头和激光雷达等数据,并将这些数据注入 32 TB 的固态存储器中。通过在试驾期间进行更多的推理和强化测试,实现 L5 级自动驾驶的过程可能会显著加快。



图 5 使用车载高性能计算平台加速 AI 算法的开发,如液体冷却装置 DYNACOR40-36(来源:Eurotech)

6 AI 会危及我们的工作吗?

社交媒体上正在讨论 AI 的进步是否会危及创意产业的工作岗位。OpenAI 推出的 DALL·E2 将自然语言需求转化为图像(见图 6)。但是,也许更令人印象深刻的是它能够真实地编辑现有图像。例如,它可以移除前景或背

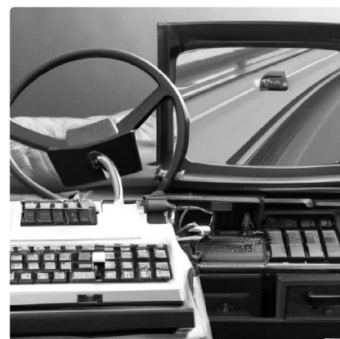


图 6 基于可信赖的 Commodore 64 驱动的 OpenAI 的 DALL·E2 生成的自动驾驶图像(一项正在进展中的工作)

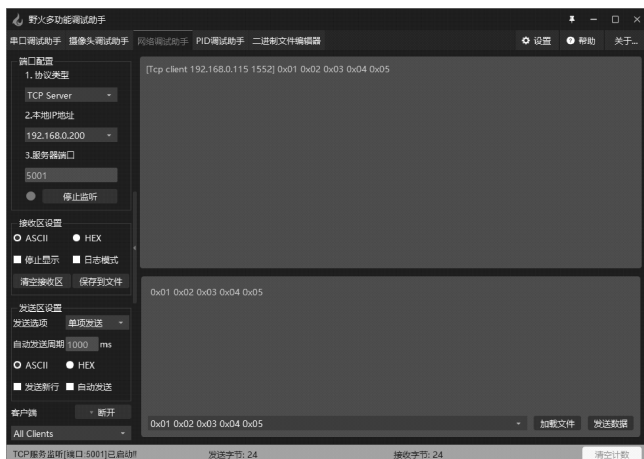


图6 网口接收信息

实验结果表明,当PC上位机通过网口向单片机主机发送数据“0x01 0x02 0x03 0x04 0x05”后,只需按下按键,单片机主机即可成功通过CAN通信向从机发送这些信息,并且将来自于从机的信息发送给PC上位机。可见,本设计成功通过FreeRTOS操作系统连接PC上位机,并与多个单片机进行数据传输。

5 结 语

针对多机通信的问题,本文通过设置FreeRTOS线程

任务,并将各个通信模块与FreeRTOS中的消息队列相结合,成功实现多机之间的通信任务,并且通过设置线程任务的方式控制多机之间的通信过程,提高了单片机的运行效率。本文设计的多机通信系统可移植性强,适用于各种需要多机配合完成的项目。本文只配置了一个任务用于与从机进行通信,如有需要还可以设置多个任务,实现与更多模块之间的通信。ME

参考文献

- [1] 张贺帅,张哲.以单片机为基础多机通信系统的设计对策分析[J].长江信息通信,2022,35(3):117-119.
- [2] 杜路泉,莫建麟,王玉晶,等.基于单片机的多机通信系统设计[J].宁德师范学院学报(自然科学版),2021,33(2):153-158.
- [3] 胡念祖.基于ARM和嵌入式Web服务器的自动售货机设计[D].西安:西安工程大学,2019.
- [4] 吴乐明.基于单片机技术的多机通信系统设计[J].中国新通信,2018,20(9):7-8.
- [5] 谭玉琪,王超,汤吉昀.基于Proteus的单片机多机通信的设计与仿真[J].电子设计工程,2019,27(3):164-168.

陈旅超(硕士研究生),主要研究方向为嵌入式运动控制。通信作者:陈旅超,2447475008@qq.com。

(责任编辑:薛士然 收稿日期:2023-05-08)

6 景中的物体。此外,如果给它一幅荷兰画家维米尔(Vermeer)的作品,AI可以扩展它的“戴珍珠耳环的女孩”,使人们对画中她所处的房间有一个可信的印象。

然而,作家(比如Elektor和其他知名媒体的编辑团队)对ChatGPT的推出感到震惊。这种AI可以以会话方式和多种语言与用户进行交互。到目前为止,关于碳化硅(SiC)MOSFET和氮化镓(GaN)晶体管优点的讨论,以及它们相对于硅MOSFET的优点,已经出现了高度精准的效果。因此,即使是小众话题,似乎也被很好地涵盖了。

虽然这个工具非常聪明,但它只知道训练之前相关话题的答案。因为它不是持续学习的,所以不会了解最新的时事或最新的K-pop乐队剧(shame)。另一个小问题

是,过了一段时间,答案似乎变得有点老套和公式化。

总之,嵌入式系统和AI越来越多地被用于创建智能、自主设备和系统中。嵌入式系统为AI算法的运行提供了硬件和软件平台,而AI算法使这些系统能够以更智能和更人性化的方式感知、分析和响应其环境。随着嵌入式系统和AI的能力不断提高,我们期待在机器人、医疗保健、交通运输等领域看到一系列令人兴奋的新应用。

本文授权来自本刊的合作伙伴Elektor媒体集团,如果希望免费订阅Elektor的英文在线内容,请访问<https://www.elektormagazine.com/>。ME

(责任编辑:芦潇静 收稿日期:2023-06-30)

IAR全面支持兆易创新基于Arm Cortex-M7内核的超高性能MCU

嵌入式开发软件和服务的全球领导者IAR与业界领先的半导体器件供应商兆易创新(GigaDevice)联合宣布,最新发布的IAR Embedded Workbench for Arm 9.40版本已全面支持兆易创新基于Arm Cortex-M7内核的超高性能MCU微控制器——GD32H737/757/759系列,为开发人员提供高效的工具链。

GD32H737/757/759系列超高性能MCU基于600MHz Arm Cortex-M7内核,凭借双发射6级流水线架构,以及支持高带宽的AXI和AHB总线接口,可以获得更高的主频及处理性能,达到1552 DMIPS和2888 CoreMarks的优异结果。GD32H7突破了MCU的性能边界,为复杂运算、多媒体技术、边缘AI等高级创新应用提供强大的算力支撑,进一步迭代拓展了兆易创新在超高性能领域的产品布局,并持续以强大供应链和高品质打造平台化优势,赋能开发者应对未来挑战。