

# 借助商业服务 破解AI科创活动的算力难题

谢作如 浙江省温州中学

**摘要:** 数据、算法和算力组成了AI时代的三大技术基础,而算力匮乏是当前中小学开展AI科创活动的主要困境之一。当真实问题涉及较复杂的AI算法和模型需要高算力支持时,购买算力服务是一种必然的选择。本文以训练一个音色转换项目为例,介绍了购买算力服务训练AI模型的真实经历,以期在算力方面为中小学AI科创活动开展提出可行的解决方案。

**关键词:** AI科创活动;算力购买

**中图分类号:** G434 **文献标识码:** A **论文编号:** 1674-2117 (2023) 07-0091-03

## ● 背景: 当算力成为AI科创活动的阻力

随着AI学习的深入,科技制作社里有学生提出想研究音频的音色转换,并在GitHub上找到了一个叫做“so-vits-svc”的开源项目。这是一种让声音听起来像指定人说出来的技术,原理是先把声音转换成声谱图,再通过深度神经网络技术训练出AI模型。利用训练好的模型,可以制作虚拟主播、游戏音效、语音合成等。

和ChatGPT一样,“so-vits-svc”也属于AIGC的一种。喜欢“泡”在B站的青少年们对这种技术非常感兴趣。遗憾的是,他们在研究这个项目时遇到了很多困难。虽然不断见招拆招,一点一点解决环境搭建的问题,但最终在训练模型时卡在了“算力”这一关。因为这

个项目的训练不支持CPU环境,没有GPU环境就无法继续前进。

## ● 分析: 算力匮乏的解决方案

对于中小学的科创教师来说,尽最大能力满足学生的学习需求是最核心的工作任务。考虑到算力匮乏问题具有普遍性,每一个开展AI教育的学校最终都会遇到,因此,笔者认真分析了算力匮乏的常见解决方案。

首先是配置算力设备。最常见的人工智能的算力设备是GPU(显卡),即英伟达(NVIDIA)的显卡。但是普通学校的办公和机房电脑,一般都不会配置GPU,即使配置了GPU,搭建环境也不容易。

其次是购买算力服务器。有条件的学校会配置算力服务器,用于AI教学和科创。这些算力服务

器会配置多款显卡,然后通过容器(Docker)技术分配给多位用户同时使用。但这些算力服务器分配的容器性能比较弱,配置的环境往往与与时俱进的开源项目不一致。

当算力不够时,购买商业算力资源才是最好的解决方案,也是高校和科研机构的常见选择。已经毕业的科技社学生在介绍他们的AI项目时,都会提到用某公司的服务器训练。2021年我们也使用过华为ModelArts的算力服务。

## ● 实例: 商业算力服务的使用

这一次,笔者选择了北京超算中心,他们的算力服务分为两种。第一种是集群资源,即云服务器。云服务器类似虚拟机,提供的是完整电脑,有系统盘、外接磁盘,可以随时恢复系统,并且拥有外网IP。如果选择的是Windows系统,可

以直接用远程桌面工具连接使用。Linux系统则通过SSH或者VNC服务。第二种是超算资源。类似大家共用一台计算机，权限隔离，用SSH命令行加载虚拟环境后，通过提交作业的方式训练模型。当然，如果对SSH不熟悉，还可以选择使用jupyter (如图1)，在网页中使用

Python代码训练模型。至于训练环境，可以自己安装，也可以求助工程师，给出软件清单即可。虽然看起来用SSH很复杂，但用pip命令安装软件本来就是基本能力，学生都能胜任。针对复杂的环境部署，超算中心也采用了多种方案，如在创建云服务器的时候，

会先告知服务器的CUDA版本。超算资源中则内置很多个CUDA版本（从9.x到11.7），用户使用“module load”命令选择相应的环境，也非常方便。笔者申请了两种服务器，分别进行了测试。相对来说云服务器最简单，有很多种算力资源可以选



图1 在超算资源上用Jupyter服务训练模型

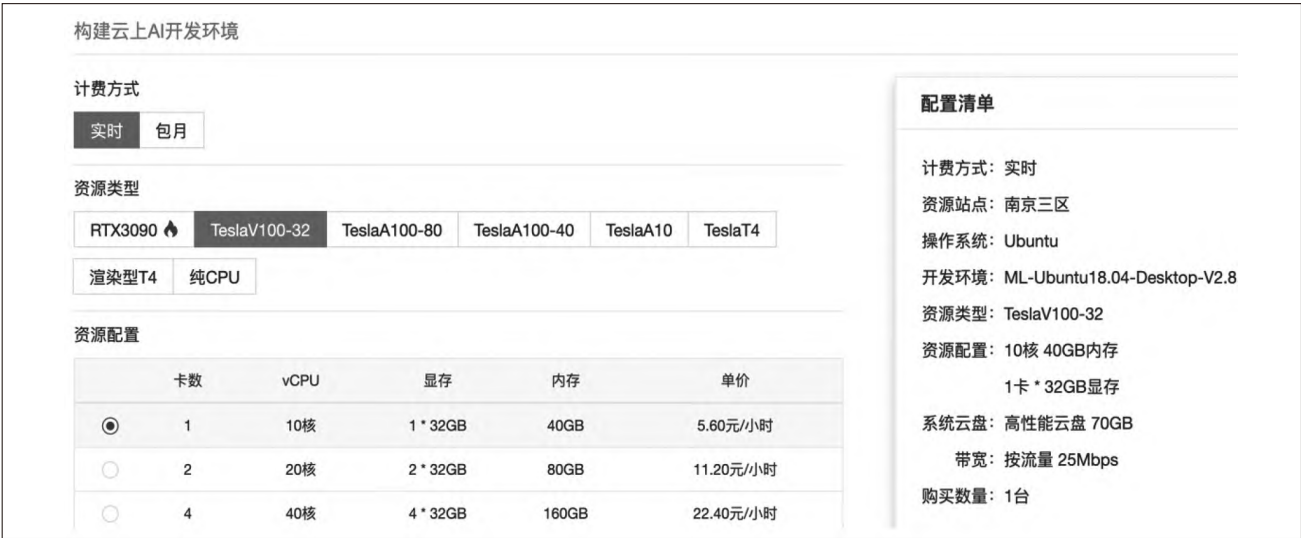


图2 在线配置GPU服务器

择。不同的算力资源对应不同版本的CUDA。笔者选择了“Tesla显卡V100-32G”(如上页图2),从配置环境到训练模型花了5个小时,花费很低,而购买这一款显卡则需要几万元。

在“so-vits-svc”项目的模型训练结束后,笔者将模型转换为ONNX格式,发现可以顺利部署在本地的普通电脑里。学生们则认为应该将模型部署在行空板上做一个有趣的变声器,于是开始了新的研究。

### ● 反思: 如何让算力的使用更加便捷

回顾这个项目的训练过程,如果超算中心的服务器上已经部署好相应的环境,那么只需要2个小时就能完成训练,需要的费用更少。而整个过程其实并没有很高的技术门槛,基本借助了Anaconda的conda命令来完成。但是对中小学而言,Anaconda也好,SSH也好,都是师生们不太熟悉的技术领域。可见,要让普通中小學生也能用购买算力的方式训练AI模型,提供算力服务的企业还需要在以下两个方面继续努力。

#### 1. 技术门槛需要继续降低

重点做好Windows系统的环境配置。例如,可以在云服务器中的桌面提供一个名为“环境配置”或者“XEdu”的快捷方式,双击即可进入虚拟环境。至于超算服务器也应该内置一个常用的AI环境,直接用“module load”命令载入即可使用。

#### 2. 内置中小学常见AI开发环境

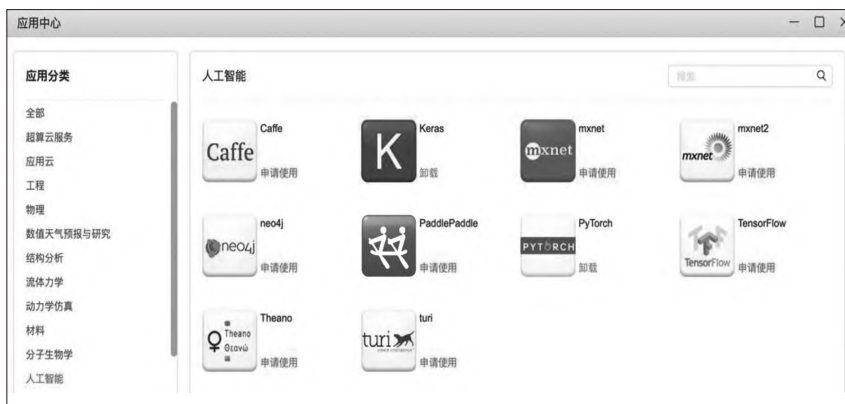


图3 超算服务器的内置应用

在超算服务器中的“应用”列表中,笔者发现“Pytorch”“Keras”“Alphafold2”“MatLab”等环境都已经内置,但还缺少中小学最常见的XEdu或者MMedu工具。经过与工程师沟通后,只要有足够的用户,他们也可以预装好XEdu,做成一个虚拟环境(如图3)。

当然,笔者认为中小学应倡导使用Linux机房。当编程语言从VB转到Python之后,在Linux上学习编程的门槛没有Windows高,何况即使在Windows下学习编程,也需要掌握conda、pip等常见命令。

### ● 小结: 让算力飞入寻常中小学校

笔者曾经在清华大学版信息科技教材的《云计算》章节写了一段话:在云计算服务的支持下,算力将如同自来水、电力一样成为公共资源。事实上,当前算力的使用依然存在技术门槛较高的问题。经过了解,北京超算中心的用户基本上来自高校科研机构,我们则是第一位中小学用户,由此可见新一代人工智能在中小学教育中并没有普遍

开展。但是,如果没有这种租借算力的经历,学生又如何真正理解算力对人工智能的重要意义呢?

实际上,在中小学开展新一代人工智能教育,最重要的研究对象是数据,即根据要解决的问题,进行针对性的采集数据,再选择开源的算法训练模型,最终部署模型并设计多模态交互。新发布的《义务教育信息科技课程标准(2022年版)》和之前的《普通高中信息技术课程标准(2017年版)》都非常重视“数据”,为人工智能的学习打下了很好的基础。因此,笔者希望提供算力服务的IT企业能看到中小学的需求,降低技术门槛,做到算力“拧开龙头”就能用,让算力匮乏不再成为AI科创活动普及的绊脚石。e