

用OpenHydra搭建本地人工智能教学平台

谢作如 浙江省温州科技高级中学

摘要: 在中小学普及人工智能教育不仅需要合适的工具和课程,还需要本地的人工智能教学平台。各种人工智能实验的开展,都需要借助特定的软硬件环境支持。OpenHydra是在这一背景下推出的国产开源项目,用于搭建一个本地的人工智能教学平台。本文分析了中小学人工智能教学平台的核心功能,并介绍了在一台算力服务器上快速搭建OpenHydra的过程,最后对OpenHydra项目提出了新的期望。

关键词: 人工智能教育; OpenHydra; 算力分割; 算力虚拟化

中图分类号: G434 **文献标识码:** A **论文编号:** 1674-2117 (2024) 07-0074-03

当数据、算法和算力被公认为是人工智能时代的三驾马车时,这就意味着生物启发式方法(搭建深度神经网络并在数据中获取智能)已经成为人工智能最重要的研究范式。中小学人工智能教育需要密切围绕“模型”开展,不能仅停留在体验和应用模型层面,还要尝试收集数据并训练、部署模型,在亲历“智能从无到有”的流程中理解模型。令人惊喜的是,随着以Keras、XEdu为代表的人工智能学习和开发工具的出现,学生只要借助少量代码甚至不写代码就能训练出人工智能模型。如果拥有合适的数据,完全可以做到解决各种真实问题。

基于这样的认识,笔者在建设学校人工智能实验室的时候,申报了一台拥有两块英伟达4090显卡的

算力服务器。经过一段时间的研究,终于在开源项目OpenHydra的支持下,搭建了局域网版本的人工智能教学平台,让一个班级的学生共享服务器算力,借助浏览器即可完成常见的人工智能实验。

● 中小学人工智能教学平台的核心功能分析

大部分人工智能学习和从业者往往是借助商业云算力来开展各种实验或者工作,但与高校不同的是,中小学人工智能教育往往在课堂上完成,对教学时间的把控要求比较高,而商业云算力并没有提供适合中小学的容器(学习环境),用户需要从零开始配置环境。最麻烦的是,在做深度学习实验时还需要在服务器和本地之间传输大量的训练数据。因此,商业云算力的方案对中小

学来说并不适合,有条件的学校需要本地搭建内置算力的教学平台,而这一平台除了能够提供教学资源和学生管理外,还能进行算力分割和学习环境管理,这是它的最核心的功能。

1. 算力分割管理

虽然借助XEdu一键安装包,在任何一台win7以上电脑上都可以开展人工智能教学,但是训练模型仅限于CPU环境,做对算力要求不高的机器学习和全连接神经网络的实验。至于基于卷积神经的图像分类和目标识别实验,没有支持CUDA(NVIDIA推出的通用并行计算架构)的GPU设备,一节课很难完成教学任务。从性价比和维护难度看,买一台算力服务器比为每一台学生电脑配置GPU的做法要

更合适。

需要强调的是,人工智能教育并不是针对少数人的高端课程,大部分学校的人工智能课是面向一个班级开展的。因此,用于教学的算力服务器需要支持30~40人的并发量。一般来说,一块24G的显卡一般可以切分为4份来使用,这样一计算,支持一个班级至少需要2~4块显卡。即便如此,40人的班级还要分组,让2~3人使用1个GPU容器,毕竟GPU设备太贵了。

2. 学习环境管理

这里的“学习环境”指的是人工智能的开发和训练环境。众所周知,搭建环境能劝退90%以上想学习人工智能的人。因此,用Docker容器预先安装好各种相关开发框架、工具和Python库,成为一种通用的选择,绝大多数的商业云算力平台都选择了内置Docker容器技术。所谓Docker容器,类似一台无桌面系统的虚拟机,借助交互API,即可在浏览器上编写代码,在远程容器上运行,再返回到浏览器上呈现。这个过程对学生用户来说是透明的,与本地操作没有区别。

目前,适合中小学生学习人工智能的环境有XEdu等。XEdu实际上是由MMEdU、BaseML、BaseNN和XEduhub等一系列Python库组成的,涵盖了Pytorch和sklearn等必备的人工智能学习框架。除此之外,Keras和PaddlePaddle也值得作为学习环境,完成一些特定的实验。为了避免Python库的版本冲突,这一平台最好

能支持多个容器。

● 用OpenHydra搭建人工智能教学平台的过程

OpenHydra是一个开源项目,旨在为中小学AI教育提供一个开箱即用的学习与实践平台。它是基于K8S(全称Kubernetes,一个用于管理容器的开源平台)开发的系统,默认内置了最新版本的XEdu。OpenHydra的开源地址:<https://github.com/openhydra/>。

1. 服务器软件准备

对于熟悉算力管理的高手来说,可以在K8S系统上直接安装OpenHydra的插件。对中小学来说推荐使用编译好的OpenHydra镜像文件来安装。只要将编译好的ISO文件烧录到U盘上,然后用这个U盘系统引导电脑启动,即可一步一步完成OpenHydra系统的安装。

2. 启动服务器并设置IP地址

全新安装OpenHydra实际上包含了安装Ubuntu(版本为server 20.04.06)和OpenHydra两个环境。安装过程类似对一台新电脑安装系统,关键操作是设置IP地址、指定磁盘和用户名。从系统引导开始计时,安装时间预计20~30分钟。需要强调的是,磁盘的数据要及时备份,安装过程中会初始化磁盘。

3. 测试教学平台

在系统安装完成后,重启电脑后在显示器上看到如图1所示的信息,说明一切准备就绪,不需要再做任何操作,就可以开始上课了。

用其他电脑打开浏览器,输入服务器地址和30001端口,如“<http://192.168.3.233:30001>”,就能看到OpnHydra的登录页面,如图2所示。

教师在上课之前,首先需要配置学生账号。“设备管理”页面将列出所有的用户登录和使用情况。

对学生而言,在OpenHydra上打开实验环境,将启动jupyter页面(如下页图3)。

● 对OpenHydra的更多功能期待

经过测试,OpenHydra的安装和配置简单,界面友好,很好地解决了算力分割的问题。但因为项目刚刚发布,还需要进一步完善。目前,OpenHydra社区汇聚了大量开源的AI开发工具和AI教学实践项目,尤其期待多方力量的加入,让这个系统能实现更多的功能。

1. 支持更多的容器环境

如下页图4所示,目前



图1



图2

OpenHydra的内置容器只有XEdu标准版, 还需要加入Keras、PaddlePaddle的容器。即使XEdu版本也应该有IOAI版本(AIO奥林匹克专用)、不同的教材版本等。甚至一些用户数量较大的课程也可以封装为一个独立的版本。这些容器可以由管理员或者教师来手动更新, 自由选择。

2. 支持在线更新课程资源

目前, OpenHydra团队提供了课程模板, 按照一定规范即可为这个项目提供标准课程。但是, 人工智能的课程往往会内置特定的数据集, 而深度学习的数据集都很多, 少的数百张图片, 多的数万张。那这些课程资源存放在哪里? 如何更新? 如果不设计在线增量更新的方式, 肯定会限制后续的发展。

3. 支持更多算力设备

众所周知, GPU设备是最昂贵的。目前, 一些国产的通用GPU设备也兼容CUDA, 因为中小学的人工智能实验不会追求最新的技术, 往往会选择一个最常用的计算框架版本, 不需要实时更新。所以, 只要做好一次适配, 就能用较长一段时间。这个适配工作意义很大, 但需要算力企业的支持。

4. 支持各种教学插件开发

中小学一旦普及人工智能教育, 各种需求就自然而然产生, 如用户通用认证、数据标注、作业评价、特定的教学活动支持等, 这些功能应该以插件的形式吸引有能力的企

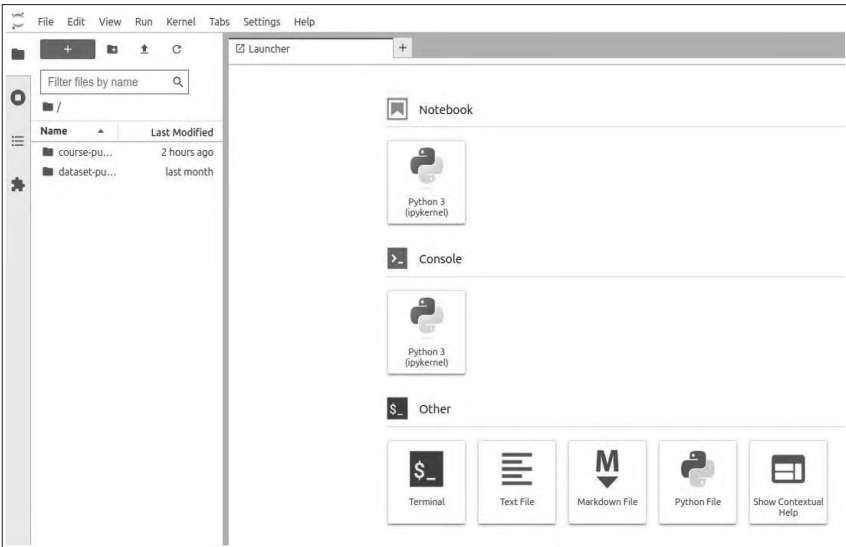


图3

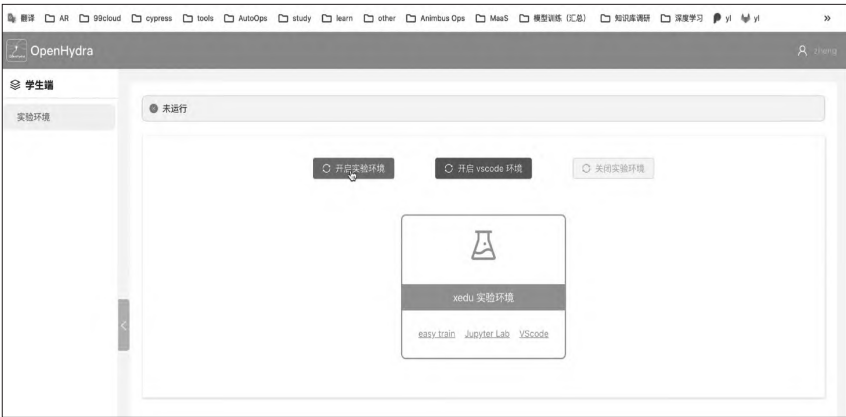


图4

业创客和教师开发。

● 总结

钟柏昌教授提出, 要开展中大型深度学习模型的训练和部署, 除了利用云服务器外, 未来开展大规模常态化人工智能课程教学, 离不开学校本地服务器的算力资源, 而这种算力资源要同时满足几十上百个模型的同步训练, 又离不开算力资源的合理分割。人工智能的发展离不开教育, 而教育需要更好的工具, OpenHydra项目不仅可以用于中小学, 而且

适用于一些规模较小的高校实验室。因此, 我们对OpenHydra项目的发展充满了期待。e