

Mind+遇上EasyTrain: 不会Python也能开发“个性化”AI项目

谢作如 浙江省温州科技高级中学

摘要: 本文介绍了不写Python代码训练AI模型的技术细节和开发AI项目的具体流程,这种AI项目开发方式对学生的编程能力要求较低,为义务教育学段实施AI教育提供了新的思路。

关键词: EasyTrain, XEduHub, AI模型训练

中图分类号: G434 **文献标识码:** A **论文编号:** 1674-2117 (2024) 13-0090-03

生活中的大量真实问题,往往需要自主训练个性化的模型来解决。这种利用自己训练的模型而开发的AI项目,笔者称之为“个性化”AI项目,毕竟现有AI模型提供的往往是“通用”能力,如任何一个图像分类模型都无法识别我家宠物和邻居家宠物的区别。之前开发“个性化”AI项目对学生的Python能力要求较高,学有余力的中学生才会试着去探究。但近几年由于相关工具的出现,训练AI模型的技术门槛越来越低,现在哪怕不会写Python代码,也能借助Mind+来开发“个性化”AI项目。

● 自主训练AI模型解决问题的流程分析

开发“个性化”AI项目大致可分为模型训练和模型应用两个环节。模型训练是指从收集数据开

始,先搭建合适的算法然后训练,最终得到具备某种能力的模型。模型应用则是将训练好的模型部署到实际场景中,解决真实问题。要让自己的AI模型“变成”一个AI应用系统,要么将模型放在某个已经编写好的程序中,要么结合某种编程语言,将模型集成到各种程序中。如图1所示,一个要解决真实问题的个性化AI项目,需要收集数据并自主训练AI模型,成为数据处理环节的

核心模块。

要实现不编写Python开发“个性化”AI项目,显然需要具备两方面的条件:其一是借助不编写Python代码的模型训练工具,收集数据并训练个性化的AI模型,其二是将这一模型整合到图形化编程工具中去,形成一个AI应用。目前,这两个条件都已经达到,前者可以借助EasyTrain结合XEdu来实现,后者可以借助XEduHub结合

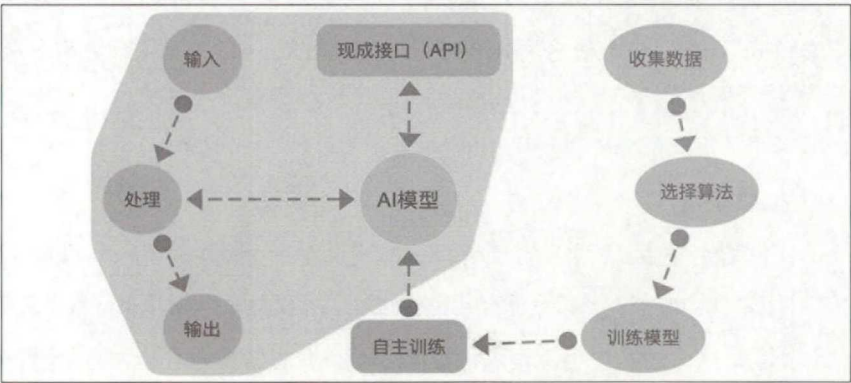


图1 一个AI项目的工作流程

Mind+来实现。

● 借助EasyTrain训练个性化AI模型

模型训练工作大致可以分为数据准备、模型搭建、模型训练、模型评估等核心环节,如图2所示。从工作流程可以看出,训练AI模型的

难度在数据准备和算法搭建环节,而训练的速度则取决于算力。只要找到合适的工具,训练常见的AI模型并不困难。

对于不熟悉Python代码的学生来说,只要安装了XEdu环境,即可使用EasyTrain来训练模型(如

图3)。EasyTrain是XEdu项目中“EasyDL”系列工具中的一种。顾名思义,它就是用来做模型训练的小工具。EasyTrain支持MMedu、BaseNN和BaseML,能够在网页点击鼠标生成模型训练代码,也能够

在网页端调用底层环境训练模型。EasyTrain位于“XEdu一键安装包”的“EasyDL”文件夹(如图4),也可以使用命令“pip install easy-xedu”单独安装。EasyTrain在训练启动时会自动检查所依赖的Python环境。如果缺少对应环境,会给出提示(如运行BaseML生成的代码需要安装BaseML,命令为“pip install BaseML”)。启动EasyTrain同样也是使用指令来完成,命令为“EasyTrain”。使用EasyTrain虽然不需要写代码,但要训练出一个AI模型不仅需要收集相关数据,还要了解机器学习的专用名词,如数据集格式、学习率(lr)、轮次(epochs)等。BaseML基于Scikit-Learn,BaseNN和MMedu基于Pytorch开发,使用的关键词和绝大多数的教程是一致的。

实际上,只要具备最基本的Python语法知识,就能借助XEdu文档编写出训练模型的代码。XEdu的系列工具训练模型的代码几乎都是公式化的,查看EasyTrain生成的代码就能看出,只需要更换数据集地址、算法名称和部分参数。当然,也可以借助大语

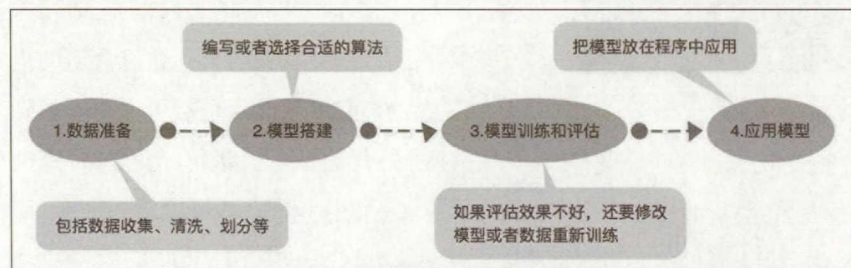


图2 人工智能模型训练的工作流程

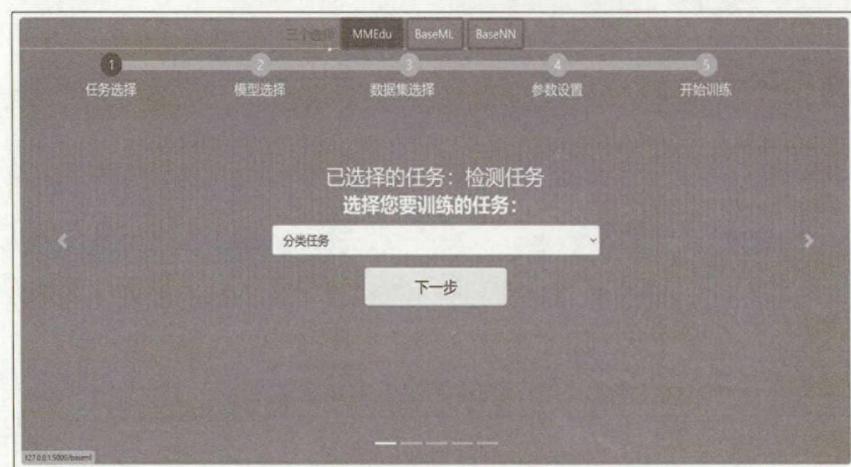


图3 EasyTrain的界面

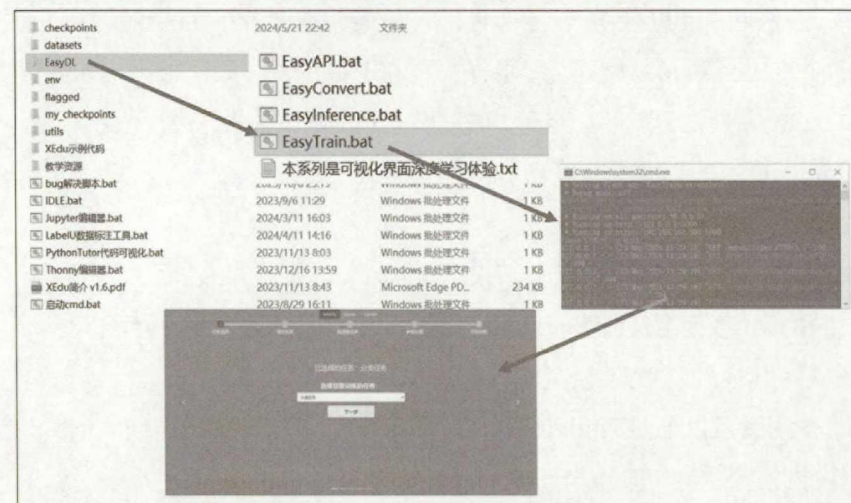


图4 EasyTrain的运行脚本位置



图5 在Mind+中推理MMEdu的模型

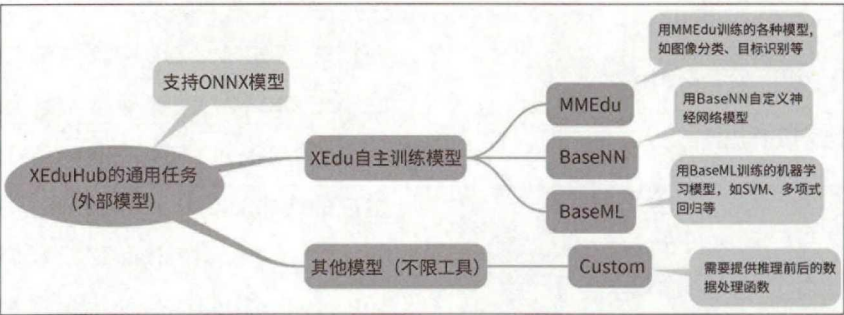


图6 XEduHub的通用任务

言模型来生成训练模型的代码。

● 在Mind+上完成个性化模型推理

几乎每一种AI开发框架都会有自己特定类型的模型，如PyTorch的模型格式为“.pth”、TensorFlow的模式格式为“.h5”等。为了让AI模型更容易转化为应用软件，不同的企业和机构相继推出了各自的推理工具（也称推理框架、部署工具）。目前，ONNX Runtime是一个应用较为广泛的推理框架，其模型格式为ONNX。EasyTrian在训练模型的最后阶段，都会提醒是否转换为ONNX格式。如果忘记转换，后期也可以用“EasyDL”中的“EasyConvert.

bat”来转换。

一般来说，训练模型的工具也会自带推理功能，如XEdu工具的XEduHub。XEduHub将自主训练的模型推理，称之为“通用任务”。“通用任务”的推理工具、语法和“预置任务”一致，也是使用Workflow模块，只要将“task”名称改为工具名称，如“mmedu”“BaseML”等即可，如图5所示。

XEduHub的通用任务不仅支持XEdu系列工具训练的模型，还支持其他的ONNX模型（如图6）。学习者可以通过ModelScope（魔搭社区）或者Hugging Face下载自己感兴趣的模型，然后通过

XEduHub推理。但前提是需要了解这个模型的输入数据格式以及输出结果数据中代表的意义。

● 结语

XEdu的核心设计理念是将AI模型训练和模型推理工作分离，降低了用AI解决问题的门槛。现在无代码训练模型的工具很多，而在大语言模型的支持下，通过合适的提示词，也能得到直接可用的模型训练代码。因此，训练常见的AI模型并非中小學生不可挑战的难题。而随着OpenHydra的出现，搭建支持算力设备的训练环境也会变得越来越简单。也许过不了多久，互联网上就会出现各种用自然语言即可完成所有操作的模型训练工具。

另外，经过本文的分析可以发现，无论是AI模型训练还是推理，对编程的要求都越来越低，几乎没有难度。但也要明确一点，要实现用AI解决真实问题，AI基础知识是必备的，如神经网络和深度学习。实际上对中小學生来说，用AI解决问题的最大难度不在算法和算力上，而在于数据——如何根据需求收集相关的数据。e