江苏省高等学校

大学生创新创业训练计划项目申报表

（创新训练项目）

|  |  |
| --- | --- |
| 推荐学院： |  |
| 项目名称： | 基于深度学习的全栈式智能错题识别平台的设计与开发 |
| 项目类型： | ☑ 重点项目  □ 一般项目  □ 校企合作基金项目 |
| 所属一级学科名称： | 工学 |
| 项目负责人： | 石智轩/陈雨桐 |
| 联系电话： | 18121656915/17586408802 |
| 指导教师： | 张成彬 |
| 联系电话： | 18921896886 |
| 申报日期： | 2023.5.12 |

江苏省教育厅 制

填表说明

一、申报表要按照要求逐项认真填写，填写内容必须实事求是表述准确严谨。空缺项要填“无”。

二、格式要求：表格中的字体采用小四号宋体，单倍行距；需签字部分由相关人员以黑色钢笔或签字笔签名。

三、项目类型为重点项目、一般项目和校企合作基金项目等。

四、项目来源：1. “A”为学生自主选题，来源于自己对课题的长期积累与兴趣；“B”为学生来源于教师科研项目选题；“C”为学生承担社会、企业委托项目选题。2. “来源项目名称”和“来源项目类别”栏限“B”和“C”的项目填写；“来源项目类别”栏填写“863项目”、“973项目”、“国家自然科学基金项目”、“省级自然科学基金项目”、“教师横向科研项目”、“企业委托项目”、“社会委托项目”以及其他项目标识。

五、表格栏高不够可增加。

六、填报者须注意页面的排版。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **项目名称** | | | | | 基于深度学习的全栈式智能错题识别平台的设计与开发 | | | | | | | | | | | | | | | |
| **项目所属**  **一级学科** | | | | | 工学 | | | | | | | **项目所属**  **二级学科** | | | | | 计算机类 | | | |
| **项目类型** | | | | | （ ）重点项目 （ ）一般项目 （ ）校企合作基金项目 | | | | | | | | | | | | | | | |
| **项目来源** | | | | | **A** | | **B** | **C** | **来源项目名称** | | | | | **来源项目类别** | | | | | | |
| √ | |  |  |  | | | | |  | | | | | | |
| **项目实施时间** | | | | | **起始时间**： 2023 年 5 月 **完成时间**： 2024 年 4 月 | | | | | | | | | | | | | | | |
| **项**  **目**  **简**  **介**  (限100字） | | 基于深度学习的全栈式智能错题识别平台是利用网络、即时通信、服务器等技术进行开发，为同学们提供一个将错题快速整理并且智能分类、排序的系统，改善传统手写誊抄错题缓慢、效率低、无法分类等问题，提高复习效果。 | | | | | | | | | | | | | | | | | | |
| **申请人或申请团队** |  | 姓名 | | | | 年级 | | | | 学号 | 所在院系/专业 | | | | | 联系电话 | | | QQ邮箱 | | |
| 主  持  人 | 石智轩 | | | | 2021 | | | | 210107012016 | 信息学院/计算机科学与技术 | | | | | 18121656915 | | | 894508104@qq.com | | |
| 陈雨彤 | | | | 2020 | | | | 200101121021 | 机械优集学院/材料成型及控制工程 | | | | | 17586408802 | | | 2044237586@qq.com | | |
| 成  员 | 王腾 | | | | 2021 | | | | 210107012018 | 信息学院/计算机科学与技术 | | | | | 19551532198 | | | 2519219745@qq.com | | |
| 冯杰 | | | | 2022 | | | | 220107121001 | 信息学院/软件工程 | | | | | 19102501034 | | | 1600311994@qq.com | | |
|  | | | |  | | | |  |  | | | | |  | | |  | | |
| **指 导 教 师** | 第一指导教师 | | | 姓名 | |  | | | | | 单位 | | | |  | | | | | |
| 年龄 | |  | | | | | 专业技术职务 | | | |  | | | | | |
| 主要成果 | | | | |  | | | | | | | | | | | | | | |
| 第二指导教师 | | | 姓名 | |  | | | | | 单位 | | | |  | | | | | |
| 年龄 | |  | | | | | 专业技术职务 | | | |  | | | | | |
| 主要成果 | | | | |  | | | | | | | | | | | | | | |
| **一、申请理由**（包括自身具备的知识条件、自己的特长、兴趣、已有的实践创新成果等）  1.知识条件  团队成员由不同专业的同学构成，成员们在校期间都已学习C语言程序设计、Java程序设计、数据库设计、计算机组成原理、操作系统等专业知识，部分成员精通Python、UML建模技术，Web前端，AI，具有软件和系统设计经验，具有一定的分析问题和解决问题的能力。成员们专业不同，背景不同，能够激发出更多的创新思想，有利于创业计划书中不同板块的撰写。  2.自己的特长  具有较多的实践活动经验，参加过多项实验、实训及实践创新活动，积极参加国际大学生程序设计竞赛（ICPC）、团体设计竞赛天梯赛、蓝桥杯、大学生数学建模竞赛、江苏省高等数学竞赛、校级大学生软件创新与设计大赛、校级大学生程序设计竞赛等且获得不错的成绩，具有较强的自主学习能力、团队协作能力和组织实施能力，并积累了一些项目研究、开发经验。算法能力较为突出，熟悉MVC程序开发，熟练掌握JavaScript、C++、Python、WebService等技术。熟练掌握Sql语言，能熟练应用MySQL、SQL Server等关系数据库。  已经具备的能力:  1）熟悉B/S框架以及前端开发技术，包括HTML、CSS、JavaScript等;  2）熟悉后端开发技术，包括Python语言、Django、Flask等框架、MySQL数据库等;  3）熟悉Linux系统的基本操作和LNMP环境的部署;  4）熟悉深度学习算法和框架，包括TensorFlow、PyTorch等;  5）熟悉机器学习算法能解决最优化及训练的数据处理;  6）熟悉python爬虫技术;  7）熟悉OCR技术和模型，如PP-OCRv2;  8）熟悉Docker技术，能够在其中运行应用程序以确保服务器的安全性;  9）具备使用UML规划系统各个功能模块流程的能力;  10）ORM框架，如Django ORM等，以提高访问效率;  11）悉企业级开发集成环境Pycharm ，Visual Studio，Eclipse ，Matlab版本控制工具GIT，命令行工具Powershell ，Terminal ，UML设计工具：Visual等;  12）备团队协作和项目管理的能力，能够与其他成员协调合作，完成项目开发;  13）备试题导入、结果反馈、历史记录、用户信息管理、错题管理、消息通知等功能的设计和实现能力;  3.兴趣  对Web开发以及服务器技术充满兴趣，热爱编程。平时通过项目设计与开发巩固了所学专业知识，对算法的编制和系统设计具有强烈的爱好，在学习上有精益求精和创新的品质  4.项目成员已有的实践创新成果  1）石智轩  2022年3月 校二等奖学金  2022年6月 校数学建模竞赛二等奖  2022年10月 校一等奖学金  2021年12月 校程序设计竞赛一等奖  2022年11月 江苏省高等数学竞赛三等奖  2022年11月 校大学生软件设计与创新大赛三等奖  2023年3月 校一等奖学金  2023年4月 第十四届蓝桥杯软件大赛C/C++程序设计本科B组省赛一等奖  2023年4月 团体设计竞赛天梯赛 团队省赛二等奖  2）陈雨桐  2021年6月 校数学建模竞赛二等奖  2021年9月 江苏省数学建模一等奖  2022年6月 校数学建模竞赛二等奖  2022年8月 江苏省高校智能机器人大赛三等奖  3）王腾  2022年11月 校级大学生程序设计竞赛二等奖  2023年3月 校三等奖学金  4）冯杰  2022年10月 国才杯初级良好  2022年11月 英语翻译竞赛二等奖  2022年12月 校二等奖学金  2022年12月 校计算机程序设计三等奖 | | | | | | | | | | | | | | | | | | | | |
| **二、项目方案**  具体内容包括：  1、项目研究背景（国内外的研究现状及研究意义、项目已有的基础，与本项目有关的研究积累和已取得的成绩，已具备的条件，尚缺少的条件及方法等）  本系统是从实际情况出发，为同学们整理错题提供一个便利、专业的错题整理平台。许多同学都热衷于整理错题，而一般都是靠手写誊抄，这种传统的整理方式非常缓慢、效率低下且无法将错误率直观地展现出来，这会后期复习的效果会产生很大的阻碍。当今社会有许多搜题软件，但是错题整理软件较为缺乏。传统的手写错题集缺少一个良好的反馈与整理能力，因此作用必定大打折扣，基于这些现实因素，我们打造了这款全栈式AI错题识别整理系统，此系统可以一次性导入大量题目，依据题目对错、类型、错误率等将它们分门别类，并存档保留。同学们可以进行在此系统上反复强化练习，确保把错题的效果发挥到最大。  2、项目研究目标及主要内容  该系统旨在为广大学生提供一个完善的错题整理系统，实时整理更新，及时反馈题目内容，且当题目知识库足够大时可实现无需提前输入答案自动判卷。改变传统手写誊抄错题的方式，改善错题整理无法智能分类和无法明确题目错误率的弊端。  3、项目创新特色概述  使用网页作为服务的终端，可以不受时间地点的限制，随时随地网上提交、练习错题。不再局限于手写错题，而是利用计算机更快更好的处理问题，帮助同学们自如方便的学习，提高复习效率。  4、项目研究技术路线  该全栈式AI智能错题平台基于B/S框架，采用阿里云ECS云服务器作为平台服务器，使用**Ubuntu 20.04**系统部署LNMP环境，拟采用基于深度学习的OCR模型PP-OCRv2 用于识别，训练数据使用互联网公开数据集及部分爬取数据集。使用TensorFlow、PyTorch等框架进行模型训练。采用Django、Flask等框架来开发API以与模型进行交互。并用Python结合MySQL数据库编写其后端，同时将代码运行于Docker中以确保服务器安全性。前端则采用HTML + CSS + JavaSciprt编写，作为用户端和后台管理端。使用UML规划系统各个功能模块的流程，使用ORM框架来提高访问效率。以实现试题导入，结果反馈，历史记录，用户信息管理，错题管理，消息通知等功能。  具体实现：   1. 训练错题识别模型：   - 数据准备：  1）数据收集：收集足够数量和多样性的数据集，包括题目和答案的图像或文本数据，以及对应的标注信息（正确/错误）。  2）数据清洗：通过去除不必要的信息和噪声，将数据集进行清洗。（去除图片中的边缘和水印，去除文本中的标点符号和空格。）  3）数据增强：为了增加数据集的多样性和数量，可以使用数据增强技术来生成更多的训练数据。（对图像进行旋转、翻转、剪裁、缩放等操作，对文本进行随机替换、删除、插入等操作。）  4）数据预处理：将图像或文本数据转换为计算机可以处理的数字表示。（对图像进行去除背景噪声、二值化、灰度化、归一化、裁剪、缩放等操作，对文本进行分词、编码、嵌入等操作。）  5）数据划分：将数据集划分为训练集、验证集和测试集。训练集用于训练模型参数，验证集用于调整超参数和评估模型性能，测试集用于最终评估模型性能。  - 特征工程：  1）特征选择：特征选择是从原始数据中选择对于预测任务有用的特征的过程。  - 过滤式方法：通过计算特征与目标变量之间的相关性或信息熵等指标，对特征进行排序或筛选。常用的指标包括卡方检验、互信息等。  - 包裹式方法：通过选择一个子集特征来训练模型，并使用交叉验证来评估子集的性能。常用的算法包括递归特征消除（Recursive Feature Elimination，RFE）等。  - 嵌入式方法：将特征选择嵌入到模型训练过程中，例如正则化方法（如L1正则化、L2正则化等）和决策树剪枝。  2）特征变换  特征变换是将原始特征转换为更有意义或更易于处理的形式的过程。常用的特征变换方法包括标准化、归一化、离散化、特征编码等。  标准化和归一化是将数据缩放到相同的尺度，以便模型可以更好地处理它们。常用的方法包括z-score标准化、min-max归一化等。离散化是将连续的数值型特征转换为离散型特征，例如将年龄分为不同的年龄组。特征编码是将类别型特征转换为数值型特征，例如使用独热编码或标签编码。  3）特征创造  特征创造是通过从原始数据中创建新特征的方式来丰富特征集的过程。常用的特征创造方法包括特征组合、特征交叉、特征拆分等。  特征组合是将两个或多个特征结合起来，例如将身高和体重结合起来计算BMI指数。特征交叉是将两个或多个特征的值相乘或相除等方式结合起来，  - 模型选择和训练：选择适合的机器学习算法，如神经网络、决策树等，将处理后的数据输入模型进行训练。  - 模型调优和测试：  2. 参数调整是指通过修改算法的超参数来调整模型的性能。这包括学习率、正则化参数、迭代次数、网络结构等。您可以使用网格搜索、随机搜索、贝叶斯优化等技术来确定最佳参数值。交叉验证技术可用于评估每个参数值的性能，以确定哪些参数值最适合模型。  3. 选择合适的评估指标：评估指标是用来度量模型性能的指标。在选择评估指标时，需要考虑业务需求和模型的应用场景。例如，对于分类问题，可以使用精度、召回率、F1分数等指标来评估模型性能。对于回归问题，可以使用均方误差（MSE）、均方根误差（RMSE）、平均绝对误差（MAE）等指标来评估模型性能。在测试模型时，建议使用多个评估指标来确保全面评估模型的性能。  4. 使用测试集：使用验证集来评估模型的性能。一旦完成了模型训练并进行了参数调整，您需要使用测试集来评估模型的性能。测试集是在模型训练和验证中从未使用过的数据，以确保模型具有良好的泛化能力。建议在测试集上使用多个评估指标来评估模型的性能。  交叉验证：交叉验证是将数据集分成多个子集的技术，以便用于模型评估和参数调整。常见的交叉验证技术包括 k 折交叉验证和留一法。在 k 折交叉验证中，数据集被分成 k 个子集，每个子集被用于一次验证，并且其余的 k-1 个子集用于训练模型。留一法是 k 折交叉验证的特殊情况，其中 k 等于数据集大小。  5. 尝试不同的模型  6．增加数据量  7. 持续监控模型：在将模型部署到生产环境之后，需要持续监控模型的性能。如果模型出现性能下降或异常情况，您需要及时采取措施，例如重新训练模型或更改参数等。  删除人为标记：  - 确认需要删除的标记：根据错题识别模型的结果，找出被识别为错题的题目，确定需要删除的标记。  - 删除标记：对于被识别为错题的题目，从数据库中删除与之相关的标记，只保留原始的题目和答案。  二、开发后端API：使用选定的后端技术栈来开发Web应用程序的后端API，使其能够接收并处理从前端发送的请求。使用Python和Django框架来开发API，使用SQLAlchemy ORM库来与MySQL数据库进行交互。  三、开发前端界面：使用选定的前端技术栈来开发Web应用程序的前端界面，以便能够提供用户友好的体验，同时能够与后端API进行通信。在Dreamweave基于HTML + CSS + JavaSciprt使用spring boot框架来构建单页面应用程序，使用JavaScript库来与后端API进行通信。  四、部署应用程序：使用选定的部署技术栈Docker来将应用程序部署到生产环境中。这可以确保应用程序在不同的环境中运行时表现一致，并且具有高可用性和可扩展性。  - 应用程序的安全性：确保应用程序具有适当的身份验证和授权机制，以防止未经授权的访问和数据泄露。  - 应用程序的性能和可伸缩性：考虑到应用程序的性能和可伸缩性，确保应用程序能够处理大量的请求和数据，并具有适当的负载平衡和缓存机制。  - 应用程序的可维护性：考虑到应用程序的可维护性，确保应用程序具有适当的日志记录和错误处理机制，并使用适当的版本控制和持续集成/持续部署（CI/CD）工具。  5、研究进度安排   |  |  | | --- | --- | | **时间安排** | **开发阶段** | | 2023.5—2023.6 | 调查问卷 | | 2023.7—2023.8 | 需求分析 | | 2023.9—2023.12 | 系统设计 | | 2024.1— 2024.2 | 系统测试 | | 2024.3 —2024.4 | 文档制作 |   6、项目组成员分工  1）石智轩：项目总负责，需求分析，系统开发  2）陈雨桐：界面总体设计，布局设计，系统开发  3）王腾：数据库设计，软件测试  4）冯杰：数据库设计，界面设计 | | | | | | | | | | | | | | | | | | | | |
| **三、学校提供条件**（包括项目开展所需的实验实训情况、配套经费、相关扶持政策等）  1、提供课题研究所需的场所，具有课题研究所需的网络设备等条件。  2、配备课题研究的指导教师和相关技术支持。  3、学校给予立项的训练计划项目经费资助。  4、提供宣传推广指导。 | | | | | | | | | | | | | | | | | | | | |
| **四、预期成果**  1、完成基于深度学习的全栈式智能错题识别平台的设计与开发;  2、系统设计说明书;  3、提交项目研究报告;  4、参加计算机学科类大赛;  5、发表课题相关论文一篇或软件著作权一项。 | | | | | | | | | | | | | | | | | | | | |
| **五、经费预算** | | | | | | | | | | | | | | | | | | | | |
| **总经费（元）** | | | **5000** | | | | | **财政拨款/企业资助（元）** | | | | |  | | | | | **学校拨款（元）** | | **5000** |
| **注：**总经费、财政拨款、学校拨款按照规定金额填写，校企合作项目企业资助金额不少于5000元。 | | | | | | | | | | | | | | | | | | | | |
| 具体包括：  1、调研、差旅费：1000元  2、用于项目研发的元器件、软硬件测试、小型硬件购置费等：2000元  3、资料购置、打印、复印、印刷等费用：500元  4、学生撰写与项目有关的论文版面费、申请专利费等：1500元 | | | | | | | | | | | | | | | | | | | | |
| **六、导师推荐意见**  签名：  年 月 日 | | | | | | | | | | | | | | | | | | | | |
| **七、院系推荐意见**  院系负责人签名： 学院盖章：  年 月 日 | | | | | | | | | | | | | | | | | | | | |
| **八、学校推荐意见：**  学校负责人签名： 学校公章  年 月 日 | | | | | | | | | | | | | | | | | | | | |