

《软件工程》课程项目 可行性研究报告

项目名称： 针对学生的个性化
评价系统

组 号： 第八组

组 长： 张睿诚

组 员： 周杨洋、杨振宇、刘冠群、
姜鑫宇、龚雨珂、陈煜斌、
陈若愚、曹灿

完成时间： 2020/10/28

研究课题：针对学生的个性化评价系统：

研究课题：针对学生的个性化评价系统：

第一章 概述

- 1.1 题目描述
- 1.2 项目背景
- 1.3 问题定义
- 1.4 软件概况
 - 1.4.1 系统的规模
 - 1.4.2 系统的目标
 - 1.4.3 现在正在使用的系统

第二章 可行性分析

- 2.1 系统的逻辑模型
 - 2.1.1 系统流程图
 - 2.1.2 数据流图
- 2.2 项目的可行性
 - 2.2.1 技术可行性
 - 2.2.2 经济可行性
 - 2.2.3 市场需求情况
 - 2.2.4 操作可行性
 - 2.2.5 社会效益分析
 - 2.2.6 风险分析

第三章 项目进度计划

- 3.1 项目工期
- 3.2 各功能的进度安排

第一章 概述

1.1 题目描述

- (1) 目的：根据学生的上课情况和学习状态，做出成绩预测以及针对每名学生个性化的评价。
- (2) 主要功能：通过学习平台或者网络爬虫获取学生上课状态和课堂讨论数据，运用深度学习自然语言处理技术，预测学生的成绩。构建心理学模型，对学生做出分类预测并做出个性化评价，实现因材施教。

1.2 项目背景

大规模管理的班级内，势必会出现学生的知识基础、学习能力明显的差异与分层，从而对老师的教学方案与进度产生困扰。因此对不同的学生制定不同的培养计划，找到其最适应的学习路线，是提高效率的最佳方式。然而因材施教的教育方式对教师提出了极高的要求，不仅需大量关注学生的学习情况，还需针对每个学生制定个性化方案。个性化评价系统将代替老师完成这些大量冗杂的工作，自动分析学生的学习状态并作出个性化评价。

1.3 问题定义

根据题目描述，问题总共有以下几点：

- 教师资源有限难以对每个学生都关心、了解到位；
- 学生缺少对自身学习状态的自知与反思，找不到薄弱点；
- 如果最后预测成绩偏低，将有助于提高学生的学习热情与动力；

1.4 软件概况

1.4.1 系统的规模

本系统为研究性项目，人力投入：9人、开发周期：11~13周。软件的使用人群为一个班级约100人的老师与学生。该项目会使用到数据库与深度学习框架，规模中等。

1.4.2 系统的目标

本系统的目标为：能够从学习平台获取到学生的学习信息，经过分析后对每个学生的成绩做出预测，并相应地给出个性化评价，帮助他们提高自身成绩。

1. 获取数据：使用爬虫从学习平台获取学生的上课情况和学习状态。
2. 分析数据：能根据收集到的信息对学生的综合状态做分析。
3. 学生分类：能根据学习状态对全体学生进行分类，对不同的类给出不同的反馈，达到个性化评价效果。
4. 成绩预测：能够对学生最后的考试成绩做出预测。

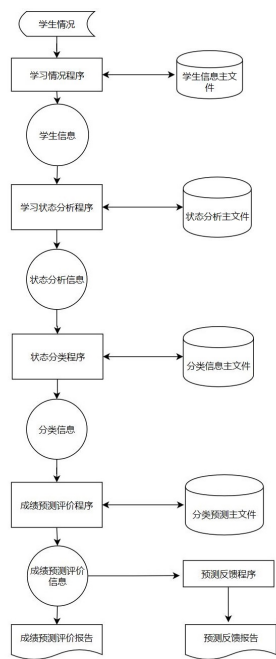
1.4.3 现在正在使用的系统

目前软件工程课程上正在使用摄像头采集听课学生的课堂表现信息，然而由于摄像头数量与角度的限制，学生的信息采集并不完整且全面。另有Moodle学习平台用于记录学生在线学习的情况。

第二章 可行性分析

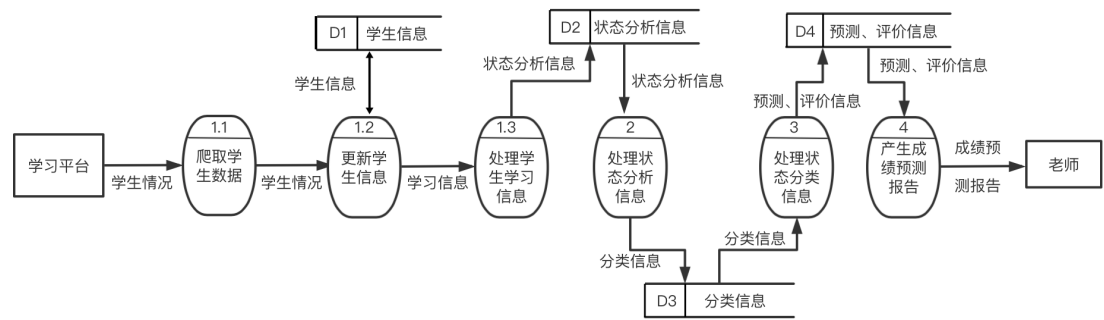
2.1 系统的逻辑模型

2.1.1 系统流程图



2.1.2 数据流图

数据流图



| 针对学生的个性化评价系统数据字典 | | | | | |
|------------------|----------|----------------------------|---------------------------------|--------------------------|-------------------------|
| | 数据名 | 描述 | 定义 | 使用特点 | 控制信息 |
| 数据源 | 学生情况 | 平台直接反映出实时的学习数据 | 学生情况=学生账户+考勤状况+作业提交状况+学习时长+更新信息 | 是输入值 | 来源于学习平台，由爬虫获取，输入到更新程序中 |
| | 学生信息 | 本程序需要使用的部分数据 | 学生信息=学生账户+考勤状况+作业提交状况+学习时长 | 是学生情况的子集，是输入值 | 是学生情况的一部分，在数据库中间滚动 |
| | 学习信息 | 能反映学生学习情况的部分数据 | 学习信息=考勤状况+作业提交状况+学习时长 | 是学生信息的子集，是输入值 | 是学生信息的一部分，由更新程序交给处理程序 |
| | 状态分析信息 | 由学习信息反映出的学习状态数据 | 状态分析信息=课堂参与度+作业完成度+学习投入度 | 反映状态分析信息的三个变量均有上下限 | 由学习信息处理后得到，会缓存并被处理程序使用 |
| | 分类信息 | 由不同的学生状态对学生进行的分类 | 分类信息=学生类别 | 由若干不同状态组成 | 由处理状态分析信息得到，会缓存并被处理程序使用 |
| | 预测、评价信息 | 对学生成绩进行预测和评价的直接信息 | 预测、评价信息=平时表现评价+期末成绩预测 | 用自然语言描述评价，成绩用数字表示 | 由处理程序输出，被报告产生程序调用 |
| | 预测、反馈信息 | 由直接信息整理得到的有效评价和预测 | 预测、反馈信息=平时表现评价+期末成绩预测 | 用自然语言描述评价，成绩用数字表示 | 由处理程序输出，由报告产生程序调用 |
| | 成绩预测评价报告 | 对期末成绩的预测 | 成绩预测评价报告=平时表现评价+期末成绩预测 | 用自然语言描述评价，成绩用数字表示 | 输出给老师 |
| | 预测、反馈信息 | 与基础成绩对比产生的数据 | 预测、反馈信息=预测+实际的差值+相差的百分比 | 相差的百分比存在上下限 | 由处理程序对比统计产生 |
| | 预测反馈报告 | 经整理生成的反馈报告 | 预测反馈报告=预测+实际的差值+相差的百分比+产生误差的原因 | 值的范围：0-实际课程 | 输出给老师 |
| 数据源分置 | 考勤状况 | 学生出勤情况 | 考勤状况=0(数字)实际课程 | 值的范围：0-实际课程 | |
| | 作业提交状况 | 学生上传作业情况 | 作业提交状况=0(数字)作业总数 | 值的范围：0-作业总数 | |
| | 学习时长 | 学生在课学习时长 | 学习时长=0(数字)课程总时长 | 值的范围：0-课程总时长 | |
| | 课堂参与度 | 由出勤情况反映的参与情况 | 课堂参与度=0(数字) | 值的范围：0-1 | |
| | 作业完成度 | 由作业提交情况反映的作业完成情况 | 作业完成度=0(数字) | 值的范围：0-1 | |
| | 学习投入度 | 由学习时长反映的投入情况 | 学习投入度=0(数字) | 值的范围：0-1 | |
| | 学生类别 | 根据学习信息把学生分为几个不同的类别 | 学生类别=高年级/低年级/差生/表现良好/表现优秀/1 | 可选项均在式中 | |
| | 期末成绩预测 | 对学生期末成绩的预测 | 期末成绩预测=0(数字)100 | 值的范围：0-100 | |
| | 学生信息 | 用于存储学生信息 | 个用于存储学生信息的存储结构 | 可输入和输出 | 由位于前后方的处理程序输入和调用 |
| | 状态分析信息 | 用于存储状态分析信息 | 个用于存储状态分析信息的存储结构 | 可输入和输出 | 由位于前后方的处理程序输入和调用 |
| 数据存储 | 分类信息 | 用于存储分类信息 | 一个用于存储分类信息的存储结构 | 可输入和输出 | 由位于前后方的处理程序输入和调用 |
| | 预测、评价信息 | 从特定网站上获取所需的重要的相关学生信息 | 一种按照一定的逻辑，自动地抽取力能回信的程序或脚本 | 由程序入口将网站信息输入程序 | 与存储信息存储结构交换数据 |
| | 更新学生信息 | 对于新录入的数据的信息进行更新操作 | 使进入程序的信息保持当前最新版本 | 根据最新的爬虫数据进行对比找出不同并存储进行交互 | 向分类信息存储结构导入数据 |
| | 处理状态分析信息 | 分析得到学生的分类信息 | 利用更新状态分析信息得到的状态信息来分析学生状态 | 由状态分析信息经分析导出学生分类信息 | 向分类信息存储结构导入数据 |
| | 处理状态分类信息 | 分析得到学生的预测、评价信息 | 利用更新状态分析信息得到的分类信息来对学生进行分类 | 由分类信息经分析导出学生预测和评价信息 | 向预测、评价信息存储结构导入数据 |
| | 产生成绩预测报告 | 综合分析数据得到学生最终的成绩预测 | 利用预测、评价信息来对学生进行最终的成绩报告 | 由预测、评价信息经分析导出成绩预测报告 | 直接输出给相关老师 |
| 处理 | 学生情况 | 学生账户+考勤状况+作业提交状况+学习时长+更新信息 | 学生信息=学生账户+考勤状况+作业提交状况+学习时长 | 是输入值 | 来源于学习平台，由爬虫获取，输入到更新程序中 |
| | 学生信息 | 本程序需要使用的部分数据 | 学生信息=学生账户+考勤状况+作业提交状况+学习时长 | 是学生情况的子集，是输入值 | 是学生情况的一部分，在数据库中间滚动 |
| | 学习信息 | 能反映学生学习情况的部分数据 | 学习信息=考勤状况+作业提交状况+学习时长 | 是学生信息的子集，是输入值 | 是学生信息的一部分，由更新程序交给处理程序 |
| | 状态分析信息 | 由学习信息反映出的学习状态数据 | 状态分析信息=课堂参与度+作业完成度+学习投入度 | 反映状态分析信息的三个变量均有上下限 | 由学习信息处理后得到，会缓存并被处理程序使用 |
| | 分类信息 | 由不同的学生状态对学生进行的分类 | 分类信息=学生类别 | 由若干不同状态组成 | 由处理状态分析信息得到，会缓存并被处理程序使用 |
| | 预测、评价信息 | 对学生成绩进行预测和评价的直接信息 | 预测、评价信息=平时表现评价+期末成绩预测 | 用自然语言描述评价，成绩用数字表示 | 由处理程序输出，被报告产生程序调用 |
| | 预测、反馈信息 | 由直接信息整理得到的有效评价和预测 | 预测、反馈信息=平时表现评价+期末成绩预测 | 用自然语言描述评价，成绩用数字表示 | 由处理程序输出，由报告产生程序调用 |
| | 成绩预测评价报告 | 对期末成绩的预测 | 成绩预测评价报告=平时表现评价+期末成绩预测 | 用自然语言描述评价，成绩用数字表示 | 输出给老师 |
| | 预测、反馈信息 | 与基础成绩对比产生的数据 | 预测、反馈信息=预测+实际的差值+相差的百分比 | 相差的百分比存在上下限 | 由处理程序对比统计产生 |
| | 预测反馈报告 | 经整理生成的反馈报告 | 预测反馈报告=预测+实际的差值+相差的百分比+产生误差的原因 | 值的范围：0-实际课程 | 输出给老师 |

2.2 项目的可行性

2.2.1 技术可行性

这个项目将结合爬虫、深度学习、心理学模型等技术，主要使用Python编写程序，可以借鉴开源的深度学习框架，因此难度合适、技术上是可行的。

功能需求：

1. 录入功能：为教师、学生提供信息录入功能，同时系统需要从数据来源获取数据并自动/手动合并到现有数据库中，这些数据将为之后的处理分析提供支撑。
2. 查询功能：为教师、学生提供查询的功能，是本系统最基本的功能。教师可以查看所有学生的信息，学生能查看自身的信息。
3. 打印功能：能打印学生的信息以及成绩预测、个性化评价报表，是本系统创新的且最重要的功能，具体实现技术可以参考学术论文或最新研究。
4. 退出功能：结束并关闭系统。

可选择的技术方案：

1. python+深度学习框架+MySQL实现：由于需要使用爬虫与深度学习等技术，使用python开发较为方便快捷；
2. 基于SSM+JSP+MySQL的网站实现：开发搭建的网站可交互性强，但对技术要求高、难度较大，且后期运营与维护费用较高；
3. 微信小程序实现：小程序的普及让开发维护的成本降低，师生也可以通过小程序更高效地查看成绩预测与报告；

局限性：

由于数据的来源包括在线学习平台、社交平台、问卷等，研究对象仅限于班内的学生，因此该系统将很难推广至其他课程或者其他学校。

数据的处理分析往往需要占用大量的时间，数据更新后无法立即查询结果。

技术说明：

- 爬虫：一段自动抓取互联网信息的程序，从互联网上抓取对于我们有价值的信息。Python 爬虫架构主要由五个部分组成，分别是调度器、URL管理器、网页下载器、网页解析器、应用程序；
- PyTorch：由Facebook人工智能研究院推出，是一个开源的Python机器学习库，用于自然语言处理等应用程序；
- MySQL：是全球最流行的开源数据库，可以经济高效地帮助用户交付高性能、可伸缩的数据库应用程序。包括Facebook、Google、Adobe、Alcatel-Lucent和Zappos在内的世界上许多规模最大、发展最快的组织都依赖MySQL来节省时间和资金，为它们的高容量网站、关键业务系统和打包软件提供动力。

2.2.2 经济可行性

投资估算

本项目的成本主要包括开发成本和运行维护成本。

（一）开发成本

开发成本主要表现为人力消耗，即开发人员的工资。我们采用代码行技术和任务分解技术相互校验来估计本项目的开发成本。

1.代码行技术

本项目的功能包括录入功能、查询功能、打印功能和退出功能，每个功能预计需要的最少代码行数、最多代码行数、最可能代码行数如下：

| | 最少代码行数 (LOC) | 最多代码行数 (LOC) | 最可能代码行数 (LOC) |
|------|--------------|--------------|---------------|
| 录入功能 | 150 | 300 | 200 |
| 查询功能 | 400 | 700 | 600 |
| 打印功能 | 250 | 600 | 450 |
| 退出功能 | 50 | 150 | 100 |
| 合计 | 850 | 1750 | 1350 |

设合计的最少代码行数为a，最多代码行数为b，最可能代码行数为m，采用下式计算所需要的代码行数的估计值：

$$L = \frac{a+4m+b}{6} = \frac{850+4 \times 1350+1750}{6} = 1650$$

按照一行代码80元人民币的平均成本，本项目所需要的人力成本约为 $1650 \times 80 = 13200$ （元）

2.任务分解技术

本项目大致可以分为可行性研究、需求分析、设计、编码和单元测试、综合测试共五个部分。每个部分需要使用的人力如下：（以人月为单位）

| 任务 | 人力 |
|---------|----|
| 可行性研究 | 1 |
| 需求分析 | 2 |
| 设计 | 5 |
| 编码和单元测试 | 4 |
| 综合测试 | 8 |
| 总计 | 20 |

按照8000元/人/月的程序员工资水平，本项目所需要的成本约为： $20 \times 8000 = 160000$ (元)

将以上两种方法计算出的开发成本相平均，可以估计本项目的开发成本约为 $\frac{(132000+160000)}{2} = 146000$ （元）

（二）运行维护成本

运行维护成本主要包括租用服务器的费用和对在使用过程中对软件进行必要的修正的费用。

1.租用服务器的费用

本软件的预期使用寿命为5年。按照每年1500元人民币的费用来估算，设年利率为5，则租用服务器的费用的净现值为

$$\sum_{i=1}^5 (\frac{1500}{(1+5\%)^i}) = 6494.22(元)$$

2.对软件进行修正的费用

预计每年对软件进行修正的费用约为1000元，则对软件进行修正的费用的净现值为：

$$\sum_{i=1}^5 \left(\frac{1000}{(1+5\%)^i} \right) = 4329.48(\text{元})$$

综上所述，软件开发的总成本为 $146000 + 6494.22 + 4329.48 = 156823.70$ （元）

经济效益

（一）定量收益

- 合同/收款金额；
- 系统为企业带来的预计开支降低；
- 系统为企业带来的预计价值增值；
- 其他如多余设备出售回收的收入等。

大约预计共有20万（元）的收益。

（二）非定量收益

无法直接用人民币表示的收益，如个性化服务的改进，信息掌握情况的改进，组织机构给外界形象的改善等。

（三）投资收益率

$$156823.70 / 200000 = 127.53$$

（四）投资回收期

大约九个半月即可收回成本。

2.2.3 市场需求情况

目前在教育领域内这种软件仍处于理论创新阶段，如果该系统能够成功上限并应对大规模的使用者，则该系统面向的各年级师生都将是潜在用户。因此未来的市场需求将会十分有前景。

2.2.4 操作可行性

该系统主要应用于学校内部，服务于高校教学，主要用户为高校的师生。该系统的逻辑模型清晰明了，软件交互简单易懂，操作上是可行的。

2.2.5 社会效益分析

社会效益方面，该系统将大大减小教师的工作压力，学生获得的教育质量也将因此受益。

2.2.6 风险分析

安全方面，有个人信息数据泄露对他人生活造成困扰的风险。应对方案：该项目采集到的数据将严格保密。

考虑法律方面的隐患，若上线运行，则需先征询用户同意采集并利用其个人信息。

数据质量方面，Moodle平台的学生信息不一定全面且完整，可以考虑从其他渠道（如问卷）获取学生信息。

该项目开发难度较大，时间紧迫。应对方案：压缩开发周期，寻找帮助。

第三章 项目进度计划

3.1 项目工期

预计项目完成所需时间：11周~13周

项目开始时间：2020年10月19日

完成项目截至时间：最后一周软件工程课

3.2 各功能的进度安排

团队共九人，将分为前端开发（3人）、后端开发（4人）、测试与文档编写（2人）三个小队同时进行开发。第一阶段五周时间，准备先补足所需的深度学习、心理学等知识体系，完成数据的爬取并分析；第二阶段六周时间，团队分别完成前后端的开发与测试、文档编写；最后一个阶段的截止时间为最后一次课的课堂演示，在此期间完善各功能、组合并完成演示demo，准备课堂展示并总结整个开发过程。

- 约两周，确定并获取数据来源，确定开发过程；
- 约五周，项目第一阶段检查；
- 约六周，项目第二阶段检查；
- 最后一周，完成软件，准备演示；