



# 基于数据挖掘技术的选课指导探索

王炼红 刘畅 周熊 许加柱 曹琳

(湖南大学 电气与信息工程学院, 湖南 长沙 410082)

**摘要:** 本文分析了当前选课制度中广泛存在的问题, 探讨了针对现有问题的改善途径, 以促进对教育大数据的分析与基于分析结果的应用。采用教育数据挖掘技术分析学生线上、线下的学习大数据, 构建了依托教学管理系统的选课指导框架。该框架集成了评估学生知识掌握状态、分析多课程关联规则、评估学生与候选课程“匹配度”等主要功能, 可为促进选课制度的改革与完善、个性化教育的落实提供参考价值。

**关键词:** 数据挖掘; 选课指导; 知识掌握; 多课程关联规则

## 1. 现行高校选课制度面临的问题

当前, 学分制已成为我国创新型人才培养模式中普遍采用的一项基本教学管理制度<sup>[1]</sup>。学分制是从选课制的基础上发展起来, 最突出的特点是允许学生选课修读。这一特点充分体现出现在学习过程中的个体自主权利与多样性, 极大地调动了学生学习的主动性<sup>[2]</sup>。但是, 经过几十年的发展, 学分制并未实现学生的完全自主选课学习。现阶段我国高校中广泛采用的选课制度为: 学校依据学生专业统一安排必修课, 学生自主选择选修课。该选课制度允许学生在掌握专业基础知识的前提下, 自主构建适合个人发展的知识结构, 避免了学生知识结构单一的弊端。但是, 这种选课制度还存在以下几点不足:

(1) 学生在学习课程之后, 得不到精准的对课程知识点掌握程度的反馈评价, 导致在选课前期对自己现有的知识基础感知模糊。

(2) 学生在选课时, 候选课程是未曾学习的课程, 这些课程的学习需要提前掌握哪些知识、对这些知识需要掌握到什么层次……这些对学生来讲都是模糊的。这可能会导致后续学习过程中, 学生因自身知识基础不能满足所选课程需要而与课程脱节。

(3) 学生缺乏合理、科学的选课指导。学生选课时跟风盲选的很多, 并不能根据自身发展需要选修课程。

这些问题很大程度上归咎于当前选课系统的不完善, 大部分高校的选课系统仅提供选课和登记成绩等基础功能, 缺乏对选课学生的后续学习情况的追踪分析, 从而也就无法提供选课辅助功能。

随着教育信息技术的发展与应用, 许多高校开始探究和实践线上、线下联合教学的教育模式, 进而积累了包含学生线上、线下学习活动的教育大数据。运用数据挖掘等技术对教育大数据进行挖掘, 可以获取丰富的有用信息。詹静<sup>[3]</sup>运用 Apriori 算法对《计算机网络基础》课程内容进行关联度分析, 探究知识之间的关联规则, 以找出影响学生成绩的深层原因, 从而指导教学改进。汪静娴<sup>[4]</sup>基于改进的 Apriori 算法分析课程与课程之间的关联性, 根据课程之间的关联性及学生的课程成绩, 可以对学生提出学业预警, 降低挂科率。徐天伟<sup>[5]</sup>考虑学生的兴趣爱好、选课记录以及学习状态三维特征, 建立了个性化课程推荐模型对学生推荐感兴趣的课程。王均霞<sup>[6]</sup>建立面向知识点的学习预警模型, 可以从学生的考试数据中挖掘知识掌握情况。牟智佳<sup>[7]</sup>基于教育测量理论建立个性化评价模型, 利用学生的答题记录诊断对应的知识水平。由此

可见, 基于数据挖掘技术, 可以评估学生的知识掌握程度, 分析课程与课程的知识关联规则, 量化学生与候选课程的匹配程度。

在此基础上, 可以开发面向学生的选课指导系统, 引导学生自主选课, 提高学生课程目标达成度, 促进毕业能力的培养。

## 2. 教育数据挖掘技术

### (一) 认知诊断理论

认知诊断理论结合了心理学、教育学和现代测量学, 从学生的认知结构出发, 评估学生作答试题的知识加工与运用过程, 最终实现对学生知识掌握水平的诊断。

DINA (Deterministic Input, Noisy “and” Gate) 模型是认知诊断理论中应用最广泛的一种基础模型:

$$P(y=1|\eta, s, g) = (1-s)^\eta g^{(1-\eta)}$$

式 1 中,  $P(y=1|\bullet)$  表示学生正确回答试题的概率。

$\eta = \prod_{k=1}^K \alpha_k^{q_k}$ ,  $\eta$  表示学生对试题的掌握状态。 $\alpha_k$  表示学生对知识点  $k$  的掌握状态,  $q_k$  表示试题是否考察知识点  $k$ 。 $s, g$  分别为试题的粗心参数与猜测参数。 $s$  表示学生在掌握试题的情况下, 因粗心导致错误作答的概率,  $g$  表示学生在未掌握试题的情况下, 通过猜测行为正确作答的概率。

基于 DINA 模型的学生知识掌握水平评估流程如下:

Step1: 由权威教师总结归纳课程知识点并标注题库中每道题目对知识点的考察情况, 得到试题-知识点关联矩阵  $Q = \{q_{jk}\}$ ,  $q_{jk}$  表示试题  $j$  是否考察知识点  $k$ ;

Step2: 获取学生试题作答详情记录, 整理得到学生试题作答矩阵  $Y = \{y_{ij}\}$ ,  $y_{ij}$  表示学生  $i$  对试题  $j$  的作答结果。

Step3: 基于马尔科夫链蒙特卡洛算法 (Markov chain Monte Carlo, MCMC) 对 DINA 模型进行求解, 得到学生的知识掌握矩阵  $\eta = \{\eta_{ik}\}$ ,  $\eta_{ik}$  表示学生  $i$  对知识点  $k$  的掌握程度。

### (二) 关联分析

关联分析是一种从大规模数据集中发现多个变量取值之间存在某种规律, 继而寻找隐藏的关联关系的技术。常用的关联分析算法主要有 Apriori 算法和 FP-growth 算法。FP-growth 是 Apriori 算法在分析大数据集时的改进算法, 可以提高分析速度。

FP-growth 算法的挖掘流程如下图 1 所示。

**基金项目:** 本文受中国高校产学研创新基金重点项目“基于线上线下多模态教育数据挖掘及应用”(2019ITA01016)、湖南省教育厅资助教改项目“采用“慕课+智慧教室”教育大数据的线上线下混合式教学研究与实践”(湘教通[2019]291号)资助。

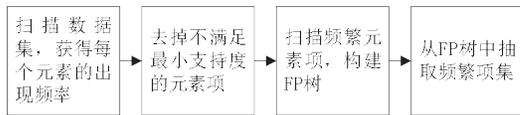


图1 FP-Growth 算法挖掘频繁模式流程

不同的课程可以看作是不同的知识点集合。知识点集合之间可能存在着相同的知识点元素, 或者一个集合中的知识点元素 A 是另一个知识点集合中知识点元素 B 的前置知识。因而, 知识点与知识点之间、课程与课程之间往往隐藏着一些关联关系。通过对知识点、课程进行关联分析, 尤其是分析候选课程知识与必修课程知识之间的关联关系, 从而可以根据学生的前置必修课程学习成果, 对学生选课指导方案做出补充与完善。

对知识点进行关联分析首先需要获取学生对学习课程的知识点的掌握程度, 然后对掌握程度进行离散, 划分为 A、B、C、D 四个等级。以 [ 知识点 id, 掌握等级 ] 作为一条记录, 扫描全体学生对所有课程知识点的掌握等级, 构建数据集。使用关联分析中的 FP-growth 算法, 基于构建的知识点 - 掌握等级数据集, 得到整个知识掌握数据的频繁项集, 根据置信度可挖掘得到选修课程知识点与必修课程知识点之间的关联规则。

对课程进行关联分析则需要对学生的课程成绩进行等级划分, 划分为 A、B、C、D 四个等级。然后以 [ 课程 id, 成绩等级 ] 作为一条记录, 扫描全体学生的所有课程成绩等级构建数据集。然后再基于 FP-growth 算法挖掘得到选修课程与必修课程之间的关联规则。

### (三) 预测算法

学生自身的某些特性是固定的, 在此基础上, 一些课程的成绩之间是具有相关性的。从这种相关性出发, 可以借助学生现有的课程成绩对未学习过的课程成绩做出预测。

预测算法需要建立训练集对预测模型进行训练, 预测模型的输入参数为现有课程的成绩, 期望输出参数为候选课程的成绩。因此, 需要用以往学生的基础课程成绩和选课课程成绩建立训练集对预测模型进行训练。训练好的预测模型应用于当前待选课学生的选课指导中去, 预测候选课程成绩作为学生与该课程的匹配度参数, 供学生参考决定是否选择该课程。BP 神经网络作为一种深度学习方法在预测算法领域取得了广泛应用, 而且数学理论已证明它具有实现任何复杂非线性映射的功能。BP 神经网络训练流程如下图 2 所示。

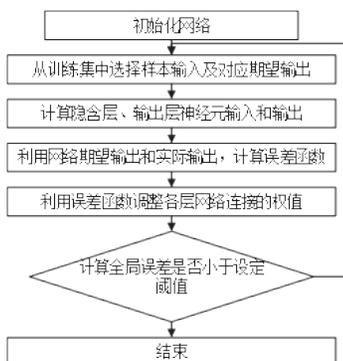


图2 BP 神经网络流程图

### 3. 基于数据挖掘的选课指导框架

基于线上、线下数据挖掘的选课指导框架如上图 3 所示。

针对线上教学管理系统, 收集学生的作业、测试、考试等学习活动的信息化数据, 采用 2.1 节中的 DINA 模型及算法对这部分数据建模分析, 挖掘评估学生的知识掌握状态, 使

学生更加清晰明确自身的知识基础。之后, 再基于关联分析挖掘课程知识之间的关联性, 构建知识点关联网络, 分析候选课程知识点与基础课程知识点之间的关联性, 为学生的选课指导提供更加细粒度的依据。

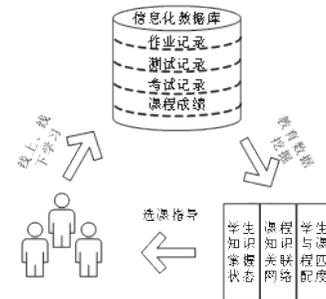


图3 基于数据挖掘的选课指导框架

线下教学管理系统记录了学生的最终课程成绩。不同于线上教学, 线下教学系统中的成绩记录虽然数据维度少, 但是记录的成绩更加公正, 也更能代表学生对该课程学习的最终成果。因此可以利用 2.2 节中的 FP-growth 算法从中挖掘学生候选课程成绩与各必修课成绩之间的关联性。继而选择与候选课程相关性强的必修课组建样本集, 对 BP 神经网络进行训练, 提高 BP 神经网络训练速度和效率, 训练好的 BP 神经网络用来预测该候选课程成绩。该成绩可作为学生与该候选课程的匹配度。匹配度越高表示学习难度越低。

教育数据挖掘的结果是形成对学生已修课程知识的评估, 对不同知识点、不同课程的关联关系的评估, 对学生与候选课程匹配程度的评估。根据评估结果, 学生既可以根据自己的知识基础进行“匹配”式选课, 也可以根据候选课程对相关基础知识的要求提前学习或复习巩固相对应的知识基础。

### 四、 效果分析

基于数据挖掘的选课指导系统如下图 4 所示。该系统应用认知诊断、关联分析、预测算法等数据挖掘技术, 分析线下教务管理系统、线上课程管理后台中的数据, 并提供两个主要功能: 诊断报告和选课建议。

如下图 4 示例, 学生张 \*\* 登录选课指导系统, 在“诊断报告”中可以查询自己已经学习过的基础课程的知识掌握评估结果。在右侧选项栏选择课程“信号与系统”, 系统生成的评估报告以雷达图的形式在下方给出。图 4 中可以看出信号与系统涉及《信号与系统基本概念》、《系统时域分析》、《信号变换》、《系统变换域分析》、《信号频域分析》、《系统频域分析》等六大知识点内容, 雷达图中分别给出学生对

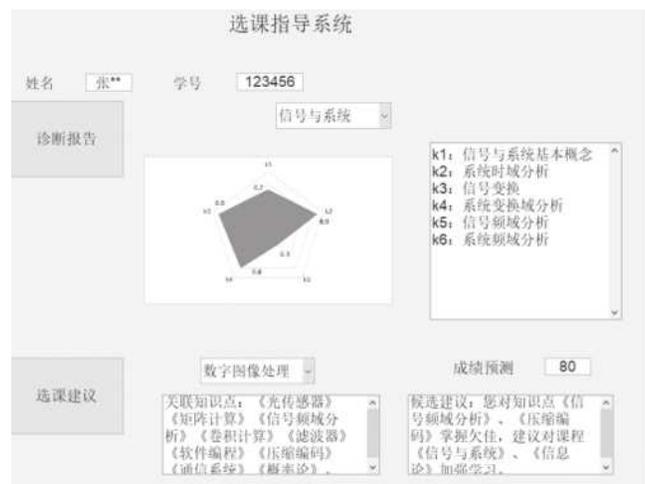


图4 基于数据挖掘的选课指导系统



这些知识点的掌握程度的量化值。“选课建议”右侧选项栏为候选课程选项，学生选择“数字图像处理”候选课程，下方显示出“数字图像处理”关联的基础知识点，如《光传感器》、《矩阵计算》、《信号频域分析》、《滤波器》、《压缩编码》等。根据该课程关联的主要知识点列表以及学生的对这些知识点掌握程度的诊断情况，给出选课建议：学生如果选择该课程，应该加强哪些知识点的学习。“成绩预测”是根据训练好的预测模型以及学生的相关基础课程成绩，对学生学习该课程的最终成绩做出预测，预测模型是基于以往的学生课程成绩记录训练的，预测结果表明了“数字图像处理”课程对学生张\*\*而言的学习难度。

根据相关课程的诊断报告，学生明确了自身的知识基础，即对课程涉及的知识点的掌握程度。根据面向候选课程的选课建议，学生可以明确自身知识基础与该候选课程的“匹配度”（成绩预测），一方面可以规避对自己来说难度较大的课程（预测成绩低的课程），另一方面也可以根据候选建议加强相关知识点及关联课程的学习。

### 5. 结束语

随着教育信息化水平的提高，面向学生的选课制度也迫切需要改进与完善。本文基于学生在线上、线下学习活动中教育大数据，探讨了基于教育数据挖掘的选课指导框架，对选课制度的改革具有参考意义与实践价值。

### 参考文献

- [1] 赵纯. 创新型人才培养视角下我国高校学分制模式研

究——基于UBC学分制案例分析[J]. 云南民族大学学报(哲学社会科学版), 2020, 37(02): 139-144.

[2] 毛正强. 对学分制条件下学生管理模式的思考[J]. 教学与管理(理论版), 2007(9): 38-39.

[3] 詹静, 范雪, 刘一帆. 基于Apriori算法的课程内容关联分析及教学策略改进[J]. 教育教学论坛, 2020, (15): 219-221.

[4] 汪静娴, 杨厚俊, 范延滨. AprioriTid改进算法在课程关联分析中的应用[J]. 工业控制计算机, 2018, 31(09): 37-38.

[5] 徐天伟, 宋雅婷, 段崇江. 基于协同过滤的个性化推荐选课系统研究[J]. 现代教育技术, 2014, 24(06): 92-98.

[6] 王均霞, 俞壮, 牟智佳, et al. 学习测评大数据支撑下面向知识点的学习预警建模与仿真[J]. 现代远程教育, 2019, (04): 28-37.

[7] 牟智佳, 李雨婷, 彭晓玲. 基于学习测评数据的个性化评价建模与工具设计研究[J]. 电化教育研究, 2019, 40(08): 96-104+113.

**作者简介:** 王炼红, 博士, 湖南大学副教授, 硕士生导师。研究方向: 信号信息处理、图像处理、数据挖掘技术、现代网络与通信技术。

(上接第2页)

### 3.3 效果分析

几天后,“作文大奖赛”成了许多学生日记和随笔的题材。他们在文中表露了自己的独特的感受。通过这些真实记录学习生活的作文,我感受到了这次活动带给他们崭新的冲击,更使我对作文评价欣赏活动中的学生参与价值进行了反思。

(1)凸显了学生作文的阅读价值。在以往的作文教学中,作文从学生笔下诞生起就以接近密封的方式来到教师的笔下,经过教师对字词句的圈点回归到学生的手中。如果不是最佳习作,它便很难再有第三位阅读者。而“作文大奖赛”活动却改变了这些作品的命运,让它有了更多的机会亮相于全体同学面前,使一个个字眼、一句句话语进入同学们的阅读视线,实现了作文被人阅读的实用价值,使读者与作者产生共鸣,形成学生间的交流。

(2)回归了学生自主学习的地位。新课程理念要求教师实现角色新的定位,即教师成为学生的促进者,从课堂教学的形式上讲就是要实现“教师搭台、学生唱戏”。在教师精心营造的“作文超市”中,“一票在手,尽得自由,”学生自主学习的权利得到了回归。他们能自由选择同学的作文进行阅读,能在阅读中享受独特的情感体验,能按照自己的喜好来评价同学的作品,能为自己喜欢的作文投上欣赏的一票。

(3)提高了学生的欣赏评价能力。欣赏评价能力是学生应该具备的基本的语文素养之一。在以往的作文讲评课中,往往是教师选一两篇典型作文进行讲评,学生的欣赏评价充其量只是在教师大意见下展开的小争鸣。现在,有了大量阅读欣赏的机会,学生在阅读了一定数量的作文后,就能在实践中进行比较区别,真正做到以参与求体验,从而领会同学

习作中的妙处,发现习作中的瑕疵。学生在文中写道:“《调皮鬼》吸引了我,我仔细地阅读起来,他真实地把同学调皮的样子细致地描写出来,让我读后忍俊不禁。”这样,学生在阅读实践中提高了欣赏评价能力,得到了真正的发展。

(4)培养了学生的写作反思意识,激发了学生的写作动力。多写多练能提高学生的作文能力,但学生在一次次不同题材的写作中经历的是一次次近乎类似的过程,很少有对自己写作活动进行反思的机会。现在,为了吸引同学的目光,得到同学的欣赏,学生开始自觉地考虑作文选材的独特,斟酌作文命题的技巧,揣摩用词造句的精妙。“评比结果出来了,我以一票之差落选了,不过,我也认识到我的作文题材不新颖,下一次我要注意选材了!”“作文大奖赛”特有的评价机制使学生有了写作反思意识,激发了学生的写作动力。

这次作文大奖赛,拓宽了学生体验、实践、交流的空间,在这自主的时间里,他们有了更多的锻炼、提高、发展的机会,在今后的作文评改教学中必将日趋完善。

### 参考文献

[1] 李莉. 以学生为主体的作文评改能力培养探究[J]. 教育科学, 2016(02).

[2] 叶树祥. 浅谈以学生为主体的作文评改方式[J]. 学业评价, 2018(1).

**作者简介:** 崔燕红(1978-),女,汉族,山东省日照市人,大学学历,中一,从事中学语文阅读,写作教学工作。