

课程绩点预测研究*

王文晶, 闫俊伢

(山西大学商务学院 信息学院, 山西 太原 030031)

摘 要: 本文通过使用人工智能和数据分析技术对学校教务系统学生成绩数据等进行挖掘评估, 对学业进行静/动态警示, 变事后处理为提前干预, 从而实现对学业预警干预的系统化管理。项目基于在研课题“应用型本科院校模型化虚拟实验教学研究”中构建的模型基础上做更深层次的教育数据挖掘, 并应用于学业预警研究中。项目拟通过对近 10 年学生成绩数据库进行研究, 运用数据挖掘 Apriori 算法对学院历年课程成绩和专业课程间的关联规则进行挖掘, 构建基于神经网络的课程绩点预测模型, 并根据预测的课程学分绩点, 提供决策支持和指导信息。该研究将帮助学生顺利完成学业; 帮助教师分层教学, 个性化指导; 有利于推动学校教学管理由“经验决策”向“数据决策”迈进。

关键词: 数据挖掘; 关联规则; Apriori 算法

中图分类号: TP391 **文献标识码:** A

0 前言

学业预警是指教学部门依据人才培养目标以及学生学分计算方法, 通过数据库中学生的成绩情况进行分析, 根据学分计算方法综合评估学生, 将学生当前的学习情况及时反馈给学生和辅导员, 有助于辅导员工作的顺利开展以及学生学业的完成。本文讨论如何根据现有课程绩点推测未上课程绩点, 可为学生选课提供建议, 能够做到提前预警, 并为教学管理者提供决策参考。人工智能时代, 将数据挖掘技术引入学业预警, 有效推动学习预警研究。本课题将运用数据挖掘中的 Apriori 算法, 挖掘出学生的考试成绩与课程之间的关联规则, 运用机器学习中的神经网络算法构建预测模型, 最终对课程成绩做一定的预测, 从而为学业预警工作提供预警信息。

随着计算机和大数据技术的发展, 学生的成绩基本实现电子存储与查询, 但是海量的成绩数据却没有深入挖掘其隐含的教学管理、教学策略价值。当前我院学生可在学生终端机随时查看自己已修课程和学分, 却无法知道自己当前的学习状态以及将来学习成绩的趋势。目前本院教学管理者, 通过教务系统中对目前成绩的评估信息, 对学生进行指导,

无法预测某课程绩点。

近年来, 将兴起的数据挖掘技术应用于学业预警, 已经成为国内高校学生管理的热点。目前有利利用数据挖掘算法对学生课程绩点进行预测, 毕业平均绩点, 还有将学业状态作为特征向量, 运用支持向量机得到非线性预测模型, 从而预测出学分绩点低的同学, 还有使用遗传优化神经网络构造成绩预测模型, 这些研究都为学业预警提供了很好的研究基础^[1]。

目前, 学业预警没有统一的方法, 不同的国内学者标准和分类也不相同, 并且学业预警在不同的国家也有很大的差异。2012 年美国教育部颁布《通过教育数据挖掘和学习分析促进教与学》, 2015 年我国国务院颁布《促进大数据发展行动纲要》, 明确提出教育教学与人才培养的战略高度应该是教育大数据^[2]。

国外对于基于数据挖掘的学习预警研究, 分别从横截面、纵贯面和技术进行。对预警指标的选择体现为横截面: Sen 等(2012)对土耳其 5000 名中学生学习成绩挖掘后, 将奖学金、平均绩点分等指标作为预警指标。纵贯面研究预警开始时间: Lee 等(2015)通过对 287 名美国大学本科生的课程数据

收稿日期: 2020-07-14

基金项目: 山西省高等学校教学改革创新项目(J2019251)

作者简介: 王文晶(1981-), 女, 山西长治人, 副教授, 研究方向为人工智能和自然语言处理。

进行挖掘,第三周开始的成绩效果与期末成绩相关。技术层体现在预警模型的选择以及改进和评估:Marbouti 等(2016)通过挖掘本科生的课堂测验、考试、作业等,使用朴素贝叶斯、人工神经网络等组合为一个模型,得到了好的预测效果^[3]。

通过文献对国内外情况进行研究对比,结果如表1所示。

表1 学习预警文献的研究数据分析

	数据	分析
文献数量		2007年至2011年较少,2012年以后论文数量逐渐的持续增长,表明数据挖掘技术在学习预警乃至教育领域,不断受到关注和重视。
文献学科分布		作为一个交叉研究领域,基于数据挖掘的学习预警研究受到了教育学、教育技术学、计算机科学等不同学科的广泛关注。
数据挖掘技术分布		多种数据挖掘方法解决学习预警问题。预测/分类方法最为常见,使用最多。
预测/分类技术分布		线性回归和逻辑回归的使用频率最高,神经网络、决策树和支持向量机的使用频率次之,此外还有少量其他技术偶见使用。

通过研究国内外关于数据挖掘的学习预警发现:近几年该方面的论文数量呈现上升趋势;数据挖掘技术采用的模型没有固定模型,比较多样化;研究点主要为模型选择、预警指标、预警起始时间。当前学业预警的主流研究方向为使用机器学习算法构造学业预警模型并预测。

1 研究内容

目前研究以课程数据成绩为基础,成绩基本上为百分制,通过研究课程之间的关联性,从而对目标课程进行预测,均未涉及到课程学分评价体系。并且,研究内容均未考虑到专业课程的整体性和关联性以及系统性,导致分析缺乏合理性和科学性,这是目前研究的一个缺口。

根据国际通行的学生质量综合评价考核体系指

标平均绩点制 GPA(grade point average, GPA),本文对构建的模型进行测试及验证,欲达到测试的准确度。一方面,教学管理者利用前导关联课程的绩点关系,预测出学生后续课程绩点,从而做出初期预警工作,让决策具有较强的说服力、科学性、合理性。学生收到预警信息后,及时改变相关联课程的学习方法和态度,从而提高目标课程绩点。另一方面,课题的相关成果,可将来作为人工智能实验室建设的教学实验资源,同时在《应用型本科院校模型化虚拟实验教学研究》教学改革项目支撑下,正在研究虚拟仿真神经网络模型,预警的相关结果将在模型中得以验证。课程成果,也可作为省级精品建设课程《数据库原理与应用》中的数据库案例和实验指导教材。

大数据兴起之后,课程绩点预测是应用型本科院校提升人才培养质量的新型教育管理方式。本课题的研究是教学实践的实际需求,又是教学研究中的一大热点。通过本课题的研究,为本院学业预警保障体系提供一种智能化、信息化策略,同时其研究成果对指导本院提升人才培养质量也具有一定的借鉴意义^[4]。

1) 本文研究数据来源于本学院近10年信息学院所有专业从入学到本科四年的期末考试成绩,所有成绩均从学校教务系统中导出,作为分析数据。对学校的教务系统中的源数据进行导出,然后将其载入数据仓库,其中一部分作为训练数据,另一部分作为验证数据。同时需对数据仓库中的数据进行准确性测试、除噪、清除错误数据。

2) 定义课程绩点计算规则。根据 GPA,结合我院学生课程成绩和综合测评的依据,经过试卷调查以及分析,最终确定本学院的课程绩点计算方法及规则。

3) 构建课程数据“画像”。学生所有专业课程都是有关联的,一门课程分数低导致后续课程会受影响。在深入研究信息学院课程体系的基础上,为分析课程之间的关联,构造知识单元、知识点、课程三个层次的专业知识有相同节点,节点之间弧线表现课程学习的先后顺序,将课程网络划分为具有共同特征的课程组。

4) 提取课程关联规则。在课程组基础上,通过 Apriori 算法对关联课程进行分析,按照支持度从高到低排序,选择几门置信度较高的课程设置为关联课程,前导课程为规则中的前置条件,预测课程设为

后置条件。这一过程 将课程之间、课程和学分绩点之间的关系模型构建出,通过分析并确定出课程间的关联系数,最终确定模型的输入和输出,即输入为什么课程的成绩和绩点,预测输出课程为什么样的后续课程成绩。

5) 构建、训练及预测学分绩点预测模型。第一,构建神经网络模型。通过 MATLAB 中的神经网络工具箱,对模型的输入层、隐含层、输出层神经元个数、参数进行计算并进行合适的参数设置;第二,训练神经网络模型。使用采集数据中的训练集,将样本的课程成绩、课程组名称作为输入数据,某些课程绩点作为输出,对神经网络进行训练,比较实际输出和期望输出差值,若误差在设定的范围内,则模型符合要求,否则调整相关参数,重复训练,直到误差符合设定值,停止迭代;第三,模型预测。使用数据集中的测试集进行测试,即给定课程组中的课程、学生、绩点,通过模型输出目标课程的绩点预测值。对比预测值和真实值之间差距,分析实验结果。

关键问题:

1) 模型建立的工具。模型建立的工具选择,既通过 Microsoft SQL Server 中的数据挖掘功能,利用原始数据构造模型,对库中已有学生进行绩点和成绩预测,也可选择 MATLAB 自带的神经网络工具箱,进行建模和数据分析。

2) 课程关联规则分析。根据本学院的课程绩点计算规则、专业知识网络和课程数据画像,运用 Apriori 算法挖掘出课程成绩和课程之间的关联性,即如何自定义评估标准,包括支持度和置信度,最终通过支持度和置信度选择先导课程和预测课程。

3) 确定模型的输入和输出项。选择将课程成绩设为输入,毕业时平均学分绩点结果设为输出,还是选择利用关联规则分析挖出课程关联规则,将课程组、成绩、绩点作为输入项目,目标课程绩点为输出向量,对比哪一种方案更加合理。

本文首先在课程内容和课程成绩基础上,需要实现的是将学生的百分制成绩数据转化为课程绩点,运用关联规则分析 Apriori 算法挖掘课程之间的关联性,从而确定模型的输入和输出;其次搭建神经网络模型,模型的训练是通过数据集中的训练集进行训练;最后,利用模型对本院课程成绩(测试集)做一定预测并分析数据,从而为学业预警工作提供相关预测信息。本课题设计方案如图 1 所示。

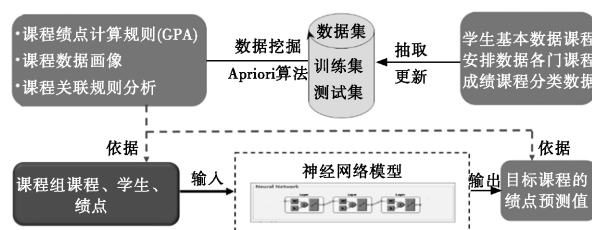


图1 设计方案图

2 结语

目前高校教学管理上较多应用百分制作为学生预警分析,本课题使用 GPA 作为学分制下评价学生成绩好坏的依据,通过创建和分析课程专业网络图,对课程之间的关联性进行分析。将课程内容和学分绩点紧密联系,结合起来分析成绩绩点,使得预测课程成绩结果更加合理和准确。

1) 课题的可行性

首先,课题的数据通过本院教务系统可获得,即实现全面综合的高质量大样本数据的获取。其次,本项目团队成员专业方向与课题一致,并且参与的项目都与本课题相关联,研究起来有更强的可操作性和专业性,通过深入的分析研究,会获得有建设性的预期成果。

2) 教研与教学管理相结合

本课题数据源真实,通过学院的成绩绩点计算规则对百分制成绩进行预处理,将课程内容和课程成绩结合进行分析,通过数据挖掘算法将本院课程间的数据关联性挖掘出,将数据挖掘技术应用于本校教学管理中。

数据的预测模型,利用机器学习神经网络搭建预测模型,该模型不但降低目标值陷入局部极小,而且也提高模型的收敛速度,最终模型计算出全局最优解,将人工智能中的机器学习(虚拟仿真)模型应用于本院的教学研究中。希望通过本课题的研究,将大数据技术应用于教学管理中,从而促进教育技术的发展。

参考文献

- [1] 吴夙慧,成颖等. K - means 算法研究综述[J]. 现代图书情报技术 2018(5): 28 - 35.
- [2] 柳玉巧. 聚类分析和关联规则技术在成绩分析中的研究及应用[D]. 武汉: 华中师范大学 2014.
- [3] 刘美玲,李熹,李永胜. 数据挖掘技术在高校教学与管理中的应用[J]. 计算机工程与设计 2010, 31(5): 1130 - 1133.

- [4] 王磊,刘清. K - Means 数据算法在大学英语四级考试成绩分析中的研究[J]. 通讯世界, 2018(3): 343 - 344.

Research on Predicting Course Grade Points

Wang Wenjing ,Yan Junya

(College of Information , Business College of Shanxi University , Taiyuan Shanxi 030031 , China)

Abstract: This paper uses artificial intelligence and data analysis technology to mine and evaluate student performance data in the school's educational administration system , including the static and dynamic warning of studies , and changing the post-event treatment to early intervention so as to realize the systematic management of early warning intervention of studies. The project is based on the model constructed in the research topic " Application-oriented Undergraduate Colleges Modeling Virtual Experimental Teaching Research " to do deeper education data mining and apply it to academic early warning research. The project intends to use the data mining Apriori algorithm to mine the association rules between the college's course scores and professional courses through the research of the student score database in the past 10 years , building a neural network-based course grade point prediction model , and based on the predicted course credits grade points , providing decision-making support and guidance information. The research will help students complete their studies smoothly , helping teachers in hierarchical teaching and personalized guidance , promoting school teaching management from " experience decision-making " to " data decision-making " .

Key words: data mining; association rules; Apriori algorithm

(上接第 68 页)

参考文献

- [1] 韩荣华. 新一代广播电视发射台自台监控系统设计与新技术应用实践 [J]. 电子世界, 2019(22) : 120 - 122.
- [2] 李云财. 对广电发射台自台监控系统的应用思考 [J]. 通信设计与应用, 2019(12) : 139 - 140.
- [3] 阮太中,孙建伟,刘平,等. 重庆市中央广播电视节目无线数字化覆盖工程自台监控系统技术方案 [J]. 视听, 2016(7) : 231 - 232.

Summary of Centralized Monitoring System of Radio and Television Transmitting Station

Zhang Guoli

(Shanxi Radio and Television Radio Management Center , Taiyuan Shanxi 030001 , China)

Abstract: At present , for the problem of the analog-digital transition of radio and television transmitting station in Shanxi Province , such as the mixed analog and digital equipment , multi-type , multi-manufacturer equipment co-exist , it is imperative to carry out centralized monitoring and monitoring of each station. Through the implementation of monitoring the transmission equipment , transmission equipment , environment and power equipment of the transmitter , the comprehensive and intuitive supervision and monitoring of all kinds of operating equipment and the environment of the engine room of the transmitter is realized , and it is sent back to the microwave station to improve the modern management level of the radio and television transmitter room.

Key words: self monitoring system; automation management; power environmental monitoring; large screen multi screen system