



# 基于综合评价模型的大学生学业成绩测量与预测

关亚琦<sup>1</sup> 卢俊香<sup>1</sup> 李海洋<sup>1</sup> 于青林<sup>1,2</sup>

(1. 西安工程大学理学院, 陕西 西安 710048;  
2. 加拿大汤姆森河大学理学院, 不列颠哥伦比亚坎卢普斯 V1S0A2)

**摘要:** 定量地了解学生的日常行为、身心健康和学业成绩之间的关系是迈向个性化教育的重要一步。与先前基于问卷调查的研究相比, 笔者收集了某高校理学院 800 名本科生的校园一卡通和综合素质测评表的数据, 选取 12 个影响因素进行主成分分析, 提取了学习勤奋度、饮食规律性和身心健康 3 个特征并与学业成绩建立非线性回归模型。同时, 利用得到的非线性回归模型对学业成绩进行预测。基于这些分析, 教育管理者可以在必要时实施针对性的干预以帮助学生提高学业成绩。

**关键词:** 校园一卡通; 综合素质测评表; 主成分分析; 非线性回归模型

**中图分类号:** TP311.13 **文献标识码:** A **文章编号:** 1003-9767 (2019) 18-001-03

## Measurement and Prediction of College Students' Academic Achievements Based on Comprehensive Evaluation Model

Guan Yaqi<sup>1</sup>, Lu Junxiang<sup>1</sup>, Li Haiyang<sup>1</sup>, Yu Qinglin<sup>1,2</sup>

(1. School of Science, Xi'an Polytechnic University, Shaanxi Xi'an 710048, China;  
2. School of Science, Thomson Rivers University, Kamloops British Columbia V1S0A2, Canada)

**Abstract:** Quantitative understanding of the relationship between students' daily behavior, physical and mental health and academic performance is an important step towards personalized education. Comparing with the previous research based on questionnaire survey, the author collects the data of 800 Undergraduates' campus card and comprehensive quality evaluation form in a college of science, chooses 12 influencing factors for principal component analysis, extracts three characteristics of study diligence, dietary regularity and physical and mental health, and establishes a non-linear regression with academic achievement. Regression model. At the same time, the obtained non-linear regression model is used to predict academic performance. Based on these analyses, education managers can implement targeted interventions when necessary to help students improve their academic performance.

**Key words:** campus card; comprehensive quality assessment form; principal component analysis; nonlinear regression model

### 0 引言

教育管理的一个重要任务是发现影响学生学业成绩的基本因素, 这对于实施教学任务, 促进学生个性化发展, 检测学生危害性行为以及干预学生的行为、情感是必要的。近几年, 随着现代化信息技术的飞速发展, 很多专家和学者通过智能手机<sup>[1]</sup>、校园一卡通<sup>[2]</sup>等前所未有的方式收集学生的生活行为、餐饮消费<sup>[3]</sup>和体育锻炼<sup>[4]</sup>的实时信息, 通过分析这些数据揭示了不同特征与学业成绩之间的相关性。文献[5]

中收集了 18 960 名本科生的校园一卡通数字记录, 以熵值提取规律性行为特征, 建立 Ranknet 模型进行成绩预测, 表明规律性是显著提高学业成绩的重要因素。在此基础上, 文献<sup>[6]</sup>采用支持向量机、朴素贝叶斯等数学方法建立预测模型, 并比较不同模型的优劣以寻找最佳模型。

然而, 目前对校园一卡通中影响学业成绩的因素研究大部分只考虑到日常行为, 并且只做出了关联性分析<sup>[7]</sup>, 极少有学者量化地考虑学生思想道德素质得分、身体素质得分和

**基金项目:** 国家自然科学基金 (项目编号: 11601410); 中国博士后科学基金 (项目编号: 2017M613169); 陕西省自然科学基金 (项目编号: 2017JM1007); 国家自然科学基金 (项目编号: 11271297); 陕西省海外百人计划资助。

**作者简介:** 关亚琦 (1993—), 女, 山西阳泉人, 硕士研究生。研究方向: 数据分析。

创新实践素质得分。在本文中,笔者通过校园一卡通和大学生综合素质测评表,收集了某高校理学院800名本科生的日常活动和德智体的数字记录,利用主成分方法提取了学习勤奋度、饮食规律性和身心健康特征,证明这三个特征与学业成绩(专业课平均成绩与创新实践成绩的总和)之间具有显著相关性,最后通过建立非线性回归预测模型来预测学生的学业成绩。特别地,思想道德素质、身体素质、创新实践素质作为影响学业成绩的重要特征第一次得到了研究。

## 1 数据来源及数据处理

### 1.1 数据来源

数据来自于西部一所工科大学的理学院800位本科生的校园一卡通和综合素质测评表,具体包含12个影响因素:进入图书馆的总次数、进入图书馆的总时长、开学前3个月进入图书馆的次数、考试前1个月进入图书馆的次数、图书馆借阅书籍的次数、借阅专业书籍的次数、消费金额、学生的思想道德素质得分、身体素质得分以及早、午、晚刷卡频次。其中,设定早、午、晚饭规律的就餐时间分别为6:00-9:00,11:00-13:00,17:00-20:00,并且统计这3个时间段内的刷卡频次。除此之外,学生的思想道德素质得分主要指学生自身的政治态度、品行表现、组织纪律以及班级小组评议得分;学生的身体素质得分主要指体育课成绩以及体测成绩。并且,将学业成绩定义为创新实践素质得分与专业课平均成绩的总和,其中创新实践素质得分主要包含学生在数学建模、“互联网+”、挑战杯比赛中取得的成绩;专业课平均成绩为2017-2018学年期末平均成绩。

### 1.2 数据处理

首先,考虑因特殊原因学生不在校的情况,删除不在校学生的数据。其次,考虑数据的可靠性和稳定性,删除学生去图书馆的总次数少于10的数据,删除学生不在规定时间内刷卡的餐饮数据。最后,对学生在图书馆及餐厅的刷卡日期进行去重处理。

## 2 模型建立及分析

### 2.1 学业成绩影响因素的相关性分析

从总样本中随机选取80%的个体作为训练集,20%作为训练集。把训练集中的12个影响因素分别用 $X_i(i=1,\dots,12)$ 表示。定义学业成绩为创新实践素质得分与专业课平均成绩的总和,通过学业成绩与影响因素之间的Pearson相关系数可知,除借阅书籍次数和图书馆借阅专业书籍次数外,其他10个自变量均与成绩存在较强的相关性,故去除借阅书籍次数和图书馆借阅专业书籍次数,借阅书籍次数与学业成绩的相关性差,可能是由于信息化得快速发展,使得学生更倾向于在网上浏览信息,查阅电子书籍。并且,在12个自变量中,学生去图书馆的次数和时长与学业成绩的相关性达0.5以上,可见学生学习的勤奋程度对学业成绩影响最大。

### 2.2 影响因素间的共线性分析

对影响学业成绩的10个自变量作共线性分析,KMO值为0.811,表示自变量间相关程度较高。Bartlett球形检验值 $p=0.000$ ,表示各自变量的独立性差,故存在共线性问题。因此,应选用主成分法对自变量进行分析。

### 2.3 影响因素的主成分分析

根据主成分列表得知,前3个主成分的特征值均大于1,累积贡献率达77.356%,表示应提取前3个主成分来代替10个自变量进行回归分析。同时,根据主成分载荷矩阵可对各主成分反映的信息进行定义。第1主成分系数的绝对值较大的变量有:进入图书馆的总次数、总时长、开学前3个月进入图书馆的次数、考试前1个月进入图书馆的次数,其对应的因素主要反映了学生去图书馆学习的勤奋度。第2主成分系数的绝对值较大的变量有早、午、晚饭刷卡频次及消费金额,其对应的因素主要反映了学生饮食的规律性。第3主成分系数的绝对值较大的变量有思想道德素质得分和身体素质得分,其对应的因素主要反映了学生的身心健康。

### 2.4 影响因素对学业成绩作用的多重回归分析

#### 2.4.1 主成分的回归方程

利用SPSS软件对3个主成分和学业成绩进行线性回归,得到函数的拟合优度 $R^2=0.529$ 。线性回归模型为:

$$Y=78.48+3.233F_1+1.03F_2+2.323F_3 \quad (1)$$

在线性回归模型的基础上,考虑非线性回归模型,根据3个主成分分别与学业成绩拟合获得的可视化相关趋势线可知,3个主成分与学业成绩之间均不满足对数及指数函数,故最终选定多项式函数对3个主成分和学业成绩进行建模。不同形式的非线性回归模型如下:

$$Y=79.48+4.7F_1+1.12F_2+2.24F_3-F_1^2 \quad (2)$$

$$Y=79.6+4.67F_1+1.11F_2+2.24F_3-F_1^2-0.13F_2^2 \quad (3)$$

$$Y=79.88+4.46F_1+1.17F_2+2.28F_3-1.63F_1^2-0.10F_2^2+0.23F_1^3 \quad (4)$$

其中, $F_1$ 、 $F_2$ 、 $F_3$ 分别表示第1、第2、第3主成分得分。模型(2)、模型(3)、模型(4)的残差分析结果分别为 $R^2=0.590$ 、 $R^2=0.591$ 和 $R^2=0.597$ ,通过比较不同非线性模型与线性模型的拟合优度,得知模型(4)的拟合效果最佳,并且模型(4)中涉及3次 $F_1$ 主成分,且系数均大于同次方的 $F_2$ 主成分与 $F_3$ 主成分,可见,学生的学习勤奋度对成绩的影响最大。

#### 2.4.2 回归方程预测效果评价

选取测试集,分别将测试集的各个影响因素代入上述线性回归方程和非线性回归方程,计算测试集的预测成绩,并且检测预测成绩和实际成绩之间的Pearson相关性。结果显示:模型(1)中预测成绩与实际成绩的Pearson相关系数为

(下转第5页)



趣味性、社交性特点的宣传体系,使垃圾分类的宣传推广更加高效<sup>[9-10]</sup>。项目实例中通过HTML5技术实现响应式网站、交互游戏及问卷表单的设计,利用HTML5技术的优点,满足垃圾分类宣传工作中朋友圈分享、后台数据收集、趣味性与科普性相结合的多种需求,对进一步开展垃圾分类宣传工作具有参考价值。

### 参考文献

- [1]Bootstrap.Bootstrap3 中文文档 [EB/OL].(2019-07-18)[2019-08-20].<https://v3.bootcss.com>.
- [2]Epub360.Epub360 官网 [EB/OL].(2019-08-05)[2019-08-20].<https://www.epub360.com/>.
- [3]W3School .HTML5 教程 [EB/OL] . (2019-8-10)[2019-08-20].<https://www.w3school.com.cn/html5/index.asp> .

(上接第2页)

0.382,模型(2)、模型(3)、模型(4)中预测成绩与实际成绩的 Pearson 相关系数分别为 0.710、0.713、0.730,显然,非线性回归方程(4)预测学业成绩的效果最佳。这表明,本文提取的学习勤奋度,饮食规律性和身心健康3个特征与学业成绩建立的非线性回归模型(4)能较好地预测学业成绩,且实际成绩与预测成绩的相关系数达 0.730。

### 3 结 语

本文主要针对学生的校园一卡通行为数据及思想道德素质、身体素质数据,为学生的创新实践成绩和专业课平均成绩提供更好的指导信息,以一卡通数据和综合素质测评表数据为基础,提取学习勤奋度、饮食规律性和身心健康特征,利用可视化图发现各个特征与学业成绩的拟合曲线,从而获得线性回归预测模型与不同的非线性回归预测模型,并对不同的模型预测结果进行比较,最终选取模型(4)对学业成绩进行预测。实验结果表明,得到的非线性回归模型(4)对学业成绩有较好的预测作用。这一结论能引导学生及时调整自我状态,改善生活规律;为教师提供了最真实、最个性化的学生特点信息,辅助信息化教学,实现因材施教。

由于本文的研究数据来源仅限于某高校的理学院,因学院自身存在个性化因素,故仅研究理学院的数据不具有足够的权威性和代表性,所得到的模型和结论不具有普遍性。作为下一步研究工作,还需要调查其他学院学生的特征及行为习惯。同时,本文中所用的数据仅限于一卡通和综合素质测评表,对学生的背景信息,如偏科情况、作息情况等无法掌握,

[4] 杨帅 .HTML5 技术在青少年科技教育中的应用探讨 [J]. 学会 ,2019(3):61-64.

[5] 扬扬 .Html5 互动广告在移动终端的用户体验探究 [J]. 艺术与设计 (理论),2018,2(11):64-66.

[6] 陈婧娴,吴祐昕 .基于 HTML5 的广告设计研究 [J]. 设计 ,2017(15):134-135.

[7] 黄永慧,陈程凯 .HTML5 在移动应用开发上的应用前景 [J]. 计算机技术与发展 ,2013,23(7):207-210.

[8] 杨臻峥,郑晓南 .利用 HTML5 页面制作工具打造科技期刊编辑全新工作方式——以《药学进展》编辑部的办刊实践为例 [J]. 科技与出版 ,2015(11):78-81.

[9] 赵荣鑫 .基于 HTML5 技术的品牌营销页面的交互设计研究 [D]. 北京 : 北京交通大学 ,2016:45.

[10] 彭杨 .Html5 可视化设计工具的易用性研究 [D]. 徐州 : 中国矿业大学 ,2016:12.

如果之后的分析中能更多地结合其他生活学习方面的数据,将大大提高学业成绩预测的准确性。

### 参考文献

[1] Wang R , Chen F , Chen Z , et al.Studentlife : assessing mental health, academic performance and behavioral trends of college students using smartphones[C]//ACM International Joint Conference on Pervasive & Ubiquitous Computing,2014.

[2] 韩伟,胡西民,付刚,等 .基于校园一卡通系统的餐饮消费数据分析 [J]. 江苏科技信息 ,2018(24):26-29.

[3] 邹志洪 .一卡通数据中学生消费行为及其成绩相关性研究 [D]. 长沙 : 湖南大学 ,2018:67.

[4]Singh A,Uijtdewilligen L,Twisk J W,et al.Physical activity and performance at school : a systematic review of the literature including a methodological quality assessment[J]. Archives of Pediatrics & Adolescent Medicine,2011,166(1):49.

[5]Cao Y,Gao J,Lian D,et al.Orderliness predicts academic performance : behavioral analysis on campus lifestyle[J]. Journal of The Royal Society Interface,2018,15(146):56.

[6] 石静 .基于数据挖掘的学生行为对学业成绩影响的研究 [D]. 武汉 : 华中师范大学 ,2017:78.

[7] 徐剑 .基于一卡通数据的消费行为与成绩的关联性研究分析 [D]. 南昌 : 南昌大学 ,2010:16.