

教学大数据监测预警平台设计与应用

李景奇¹ 卞艺杰^{1,2} 黄波³

¹(河海大学商学院 江苏 南京 211100)

²(江苏省“世界水谷”与水生态文明协同创新中心 江苏 南京 211100)

³(河海大学教务处 江苏 南京 211100)

摘要 发展性评价是我国教学评价模式改革的重要方向,但发展性评价存在组织困难、数据难以收集、反馈困难等问题。大数据技术为发展性评价提供新的技术推动力量。通过教学发展性评价中的大数据需求分析,提出教学大数据监测预警的概念,并探讨其内涵及意义。设计教学大数据监测预警平台,构建其总体架构、功能结构、实体画像和预警模型,并对其应用机制、硬件架构、模型应用进行分析和实证。该技术方案对大数据在教学过程中的应用具有积极的借鉴作用。

关键词 发展性评价 教学大数据 监测预警 实体画像

中图分类号 TP315 **文献标识码** A **DOI**:10.3969/j.issn.1000-386x.2019.01.020

DESIGN AND APPLICATION OF MONITORING AND WARNING PLATFORM FOR BIG DATA IN TEACHING

Li Jingqi¹ Bian Yijie^{1,2} Huang Bo³

¹(Business School, Hohai University, Nanjing 211100, Jiangsu, China)

²(Jiangsu Provincial Collaborative Innovation Center of World Water Valley and Water Ecological Civilization, Nanjing 211100, Jiangsu, China)

³(Office of Academic Affairs, Hohai University, Nanjing 211100, Jiangsu, China)

Abstract Developmental evaluation is an important direction for the reform of teaching evaluation model in China, but it is difficult to organize, to collect data, and to feedback. Big data provides a new technological driving force for developmental evaluation. This paper put forward the concept of monitoring and warning of teaching big data, and discussed its connotation and significance by analyzing the demands of big data in teaching development evaluation. We designed the monitoring and warning platform for big data in teaching, and constructed its overall architecture, function structure, entity portraits and warning models. We analyzed and verified its application mechanism, hardware architecture and model application. The technology scheme has a positive reference for the application of big data in teaching process.

Keywords Developmental evaluation Teaching big data Monitoring and warning Entity portrait

0 引言

随着大数据技术的发展,教育评价对大数据的需求日益强烈。我国教育评价存在目标比较狭窄、方法相对陈旧、主体比较单一、结果的呈现过于简单等问

题,需要进一步改进教育评价技术,科学解读教育评价结果,为学生和教师提供信息反馈,提供发展性评价^[1]。“评价最重要的目的不是证明,而是改进”^[2-3]的理念逐渐深入人心,发展性评价已经成为教学评价模式改革的重要方向。但如何将发展性评价理念应用到实际的教学中,存在着几个难以解决的问题:

第一,发展性评价组织困难。发展性评价需要在教学过程中不断地给予测试、监督,及时发现问题。但在高校教学中,由于教师科研压力较重,平时的作业和测试很少,成绩更多地取决于期末考试。

第二,发展性评价数据难以收集。美国心理学家桑代克认为:凡物的存在必有其数量。发展性评价更需要量,而且是非常大的量。这种“大”表现在时空两个层面上。在时间上,所需要的过程数据非常多。在空间上,所需要收集的数据范围非常广^[4]。

第三,发展性评价反馈困难。发展性评价要发挥效果,必须及时把评价结果反馈给评价对象。但是,传统技术条件下,这一点很难做到^[5]。

以上种种问题,无不是发展性评价推广应用所必须面对的问题。针对以上问题,本文提出基于教学大数据平台,建立监测预警系统,通过数据的自动收集和自动评价,并通过积极反馈,达到发展性评价所需要数据要求和过程性评价要求。

1 教学大数据监测预警的内涵及意义

1.1 内涵

教学大数据监测预警是通过大数据平台,收集分散在智慧校园中各个业务系统中教学相关数据,通过数据处理,向教学管理人员和教师以可视化的界面展示教学过程监测情况,然后通过数据挖掘和学习分析手段,对数据进一步建模分析,向教学管理人员、教师及学生提供预警信息。

1.2 意义

(1) 有利于发展性教学评价的推广和应用。教学大数据监测预警平台为发展性评价提供了数据处理的基础。借助大数据集成环境,收集数据,然后利用分布式大数据处理平台,对各类教学数据进行处理,然后对各类数据根据需求进行分类和转换,为监测预警提供分析平台。

(2) 有利于提高教学过程管理水平。教学大数据监测预警为教学监测提供更为快捷的技术手段。传统上教学过程监测耗时耗力,组织困难。而利用大数据手段,可以自动化地采集教学过程数据对教学过程进行监测。

(3) 有利于提高教学反馈效率。教学大数据监测预警平台为教学预警提供数据建模和挖掘的平台,通过数据分析和建模,主要是要发现其中的异常信息。相对于监测主体,异常信息占总体比例一般较低。因此只对异常信息进行预警,使相关人员及早了解、及早

预防,同时也减轻了教学管理人员和教师的反馈负担,提高了反馈效率。

2 平台设计

2.1 总体架构设计

教学大数据监测预警平台包括三个部分:教学大数据平台、教学监测系统、教学预警系统。教学大数据平台主要功能是实现学生网络学习行为数据的收集以及各类教学业务数据的共享。教学监测系统是对大数据平台的数据的进一步加工,根据业务需求和监测需要对数据进行联合和萃取,结果以可视化的形式展示出来。教学预警系统则是在监测数据的基础上,进一步根据预警模型,进行数据挖掘和分析,重点发现异常数据,及时提醒。具体如图1所示。

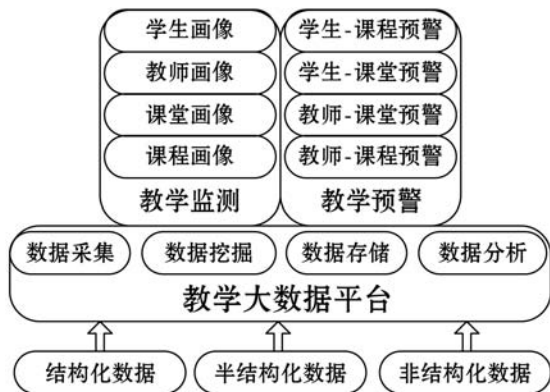


图1 教学大数据监测预警平台总体架构图

教学大数据平台对结构化数据、半结构化数据和非结构化数据进行采集,然后对数据进行挖掘处理,对处理结果进行存储,并对数据进行分析。

教学监测系统的主要展现方式包括学生画像、教师画像、课堂画像、课程画像。学生画像即学生的标签化,是依据学生的基本属性和学习过程特征而抽象出的标签化学生模型^[6]。学生画像的标签主要包括学生的基本信息、学业状态、学习风格。教师画像的标签主要包括教师的基本信息、科研状态、教学状态、教学风格。课程画像围绕课程教学而产生相关教育数据,包括课程基本信息、课程成员、课程资源、课程作业、师生交互行为、课程考核等数据^[7]。课堂画像主要包括课堂的基本信息、学生出勤情况、网络课堂的交流情况、学生的学习成绩。

教学预警主要是通过对监测指标的分析,根据阈值显示警示信息。教学预警主要是针对教师和学生在学习过程的异常情况,实现形式主要包括学生-课程预警、学生-课堂预警、教师-课程预警、教师-课堂预警。预警过程包括三个环节:第一,异常信息发现。通过预

警模型,发现异常信息。第二,对异常信息以信号灯的形式表现出来。信号灯一般可设置三种状态:即红黄绿。红灯代表严重警示,黄灯代表一般警示,绿灯代表正常。如果需要也可以增加或减少信号灯颜色。红灯预警是教师和管理人员重点关注的预警信号。第三,异常信息的处理与反馈。异常信息应及时反馈出去,通过短信、邮件等形式通知到管理人员和教师,使其能及早获得异常信息,以便做出对策。

2.2 功能结构设计

教学大数据监测预警平台主要包括三个功能模块:系统管理、教学监测和教学预警。系统管理负责对用户、数据及各类模型的管理。标签管理主要是对实体画像所要展示的标签进行定义,对标签的涵义及相关建议进行说明。模型管理主要是对数据挖掘和分析的模型进行登记,记录其方法、过程、数据集类型、结果及应用情况。教学监测主要是对教学相关的要素进行监测,包括学生画像、教师画像、课堂画像、课程画像等,如图2所示。

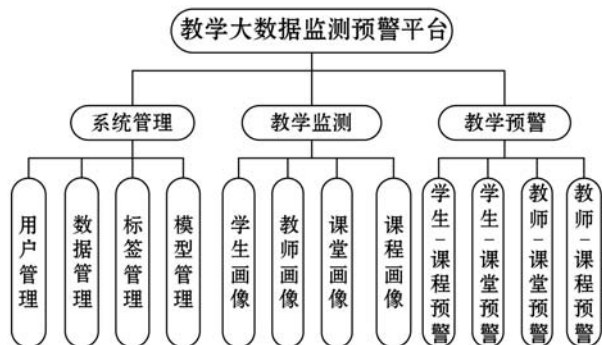


图2 教学大数据监测预警平台功能结构图

画像的展示分为两个层次,首先是以列表和图表的形式展示整体的按标签统计的信息。其次根据用户需求,点击相应记录,查看个体更为具体的画像信息。

学生预警包括学生-课程预警和学生-课堂预警。学生-课程预警是指在课程上课之前,教师获得的学生学习该课程存在的风险警示信息。学生-课堂预警则是在课堂教学中,教师根据课堂教学信息,及时预判学生的学习情况。教师预警则是对教师教学情况的预警,指标包括其历史教学情况和当前所带的所有课程教学情况,以支持教学管理人员对教师的教学评价,它又可以包括教师-课堂预警和教师-课程预警。教师-课堂预警是对课堂教学情况的预警,通过课堂教学指标数据,及时发现问题,支持教学管理人员对教师课堂教学的监督。教师-课程预警是对整体课程评价情况的预警,主要是对课程评价指标进行细化分析,及时发现课程中存在问题,为是否开课提供决策支持。预警同样也包括两个层次,分别以列表形式和图表形式展现,

然后用户能根据需要进行进一步查看详细预警信息。同时,如果用户根据实际判断预警错误,也能对预警信息进行修复,使预警信息不再呈现。

2.3 实体画像

学生画像从五个维度来呈现学生学习特点,这五个维度分别包括学习风格、学习成果、学习特长、学习基础、学习成绩。而每个维度按照相应指标得分,分别用较为通俗的标签来表示学生的学习特点。如学习风格,根据学生的一天中学习效率的高低,比如上午学习效率,则可归类为百灵鸟标签类型,如果晚上学习效率高,则可归类为猫头鹰标签类型等,如图3所示。总之,根据这五个维度的得分情况,归纳出多种标签,通俗、形象地呈现学生的主要学习特点。而学生画像可以供教学管理人员、教师、学生查看。

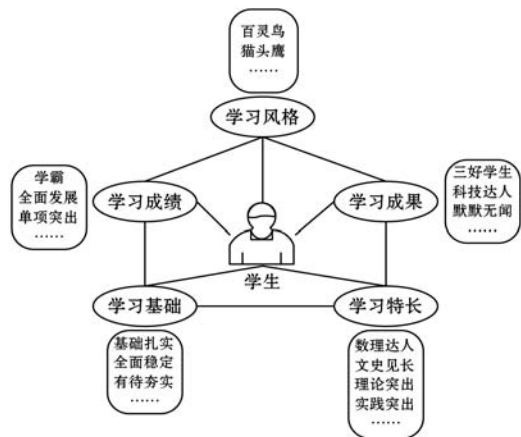


图3 学生画像

教师画像从五个维度来呈现教师特点,这五个维度分别包括学术成果、教学成果、任课门数、学生评价、学生成绩。而每个维度按照相应指标得分,分别用较为通俗的标签来表示教师的教学能力特点。如学生评价维度,可以根据学生评价的得分,分为超级巨星、褒贬不一、漠不关心来形容学生对教师教学的态度和看法,如图4所示。总之,根据这五个维度的得分情况,归纳出多种标签,通俗、形象地呈现教师多维度的教学特点。

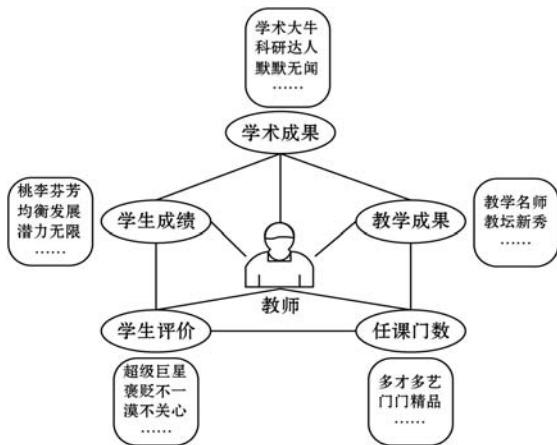


图4 教师画像

课程画像从课程荣誉、课程类型、任课教师、选修人数、学生成绩五个维度来刻画课程的基本情况。如课程类型,可分为理论抽象、应用实践、人文社科、数理自然,从这些标签类别可以看出课程的基本类型,从而也可以直观性地得出课程的教学情况,如图5所示。

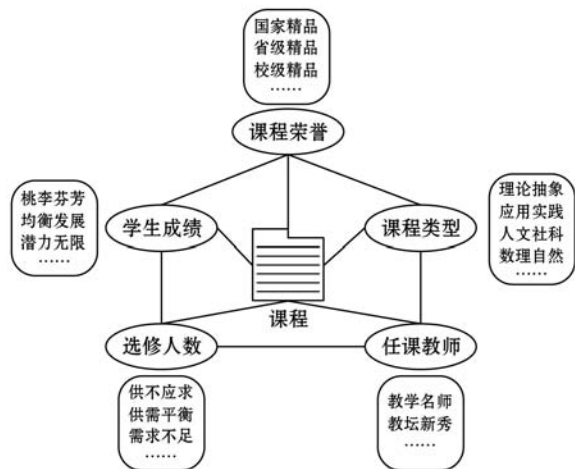


图5 课程画像

课堂画像从课程荣誉、出勤率、任课教师、课堂人数、学生成绩五个维度来刻画课堂的基本情况。如出勤率,标签可分为场场爆满、略有空位、门可罗雀,从这些标签类别可以通俗易懂的标签语言看出课堂的基本情况,如图6所示。

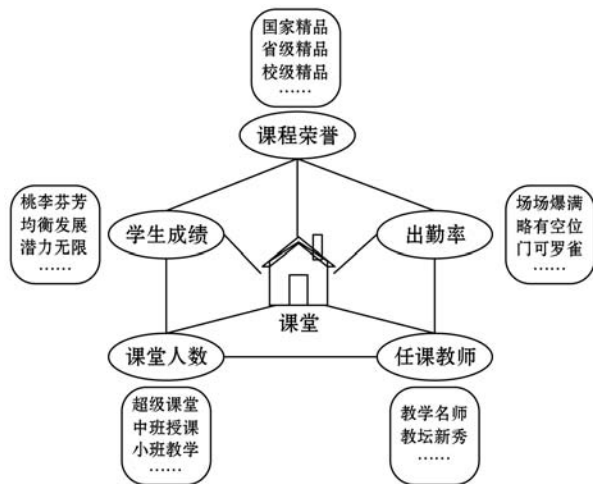


图6 课堂画像

2.4 预警设计

预警信息本身包括两个部分,一个是预警对象,另一个是预警结果。预警对象如图7所示,对象包括学生、教师、课堂、课程,结果包括红绿灯信号信息及其解释。而对对象间的相互作用的预警信息包括学生-课堂预警、学生-课程预警、教师-课堂预警、教师-课程预警四类。学生预警自动反馈给学生,教师预警自动反馈给教师,学生预警教师可查询,但不自动反馈。学生与教师预警都可以被教学管理员主动查询,但不会自动反馈。

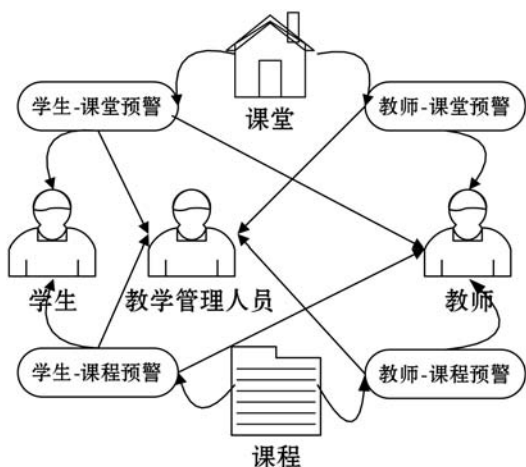


图7 预警信息功能结构图

学生-课堂预警:根据学生出勤率、作业和测试情况,设定综合得分阈值,凡低于该阈值,自动报警到该学生邮箱或手机、微信等信息接收终端。并提供给教师和教学管理人员查询该预警信息。

学生-课程预警:根据往年课程学生成绩,通过多元回归模型,计算学生该课程的预测成绩,低于60分,则自动报警给该学生,并提供给教师和教学管理人员查询该预警信息。

教师-课程预警:根据选课情况、学生评教和学生成绩情况,自动将预警信息反馈给教师,并为教学管理人员提供预警信息查询。

教师-课堂预警:根据出勤情况和网络教学作业和测试批改信息,设定综合阈值,低于该阈值,则自动反馈报警信息至教师,同时为教学管理者提供预警信息查询。

2.5 预警模型

预警模型是用于产生预警信息的数据处理模型,在本方案中,预警模型采用通用的方式表示,预测值都表示为各变量的加权之和:

$$y = \sum_{i=1}^n a_i x_i \quad (1)$$

式中: y 是计算所得的预测分值, x_i 是第 i 个变量, a_i 为变量 x_i 的系数或权重。该预警模型易于在模型库中表示,只需要记录系数(或权重)和变量值即可实现。而在具体应用中,该模型又有两种模型生成方法,分别是多元线性回归方法和综合指标加权方法。

(1) 多元线性回归方法 多元线性回归方法在成绩预测中多有应用^[8]。本方案中多元线性回归方法主要用于学生-课程预警。学生-课程预警首先获取该课程的所有前置课程(一般可选取上一学期或学年的所有相关课程),然后求它们的相关性,并取出所有相关性强的前置课程。然后通过多元线性回归算法,计

算回归系数,则式(1)变为如下所示预警模型:

$$P_c = \sum_{i=1}^n P_{ci} R_{ci} \quad (2)$$

式中: P_c 是指该课程 C 的预测成绩, P_{ci} 是指前置课程 C_i 的成绩, R_{ci} 是前置课程 C_i 的回归系数。

(2) 综合指标加权方法 综合指标加权方法原理如下:首先获取各预警指标的数值,如果是成绩数据则直接引用。非成绩数据采用极值法进行标准化,然后乘以 100,使其得分在 $[0, 100]$ 区间内,计算方法如下:

$$P_{ci} = \frac{V_{ci}}{\text{MAX}(V_i)} \times 100 \quad (3)$$

式中: $\text{MAX}(V_i)$ 为第 i 项指标最大值, V_{ci} 是该预警对象第 i 项指标的具体值, P_{ci} 是预警对象的非成绩指标项得分,即式(1)中的变量 x_i 。变量 a_i 在这里用权重 W_{ci} 代替,则式(1)可由下式表示:

$$P_c = \sum_{i=1}^n P_{ci} W_{ci} \quad (4)$$

式中: P_c 是指该对象的预测分值, P_{ci} 是指各指标项得分, W_{ci} 各指标项的权重。权重的确定方法有两种,一是根据经验或者邀请专家进行打分,确定权重;二是通过对历史数据的机器学习取得。方法二的具体算法是首先给出初始权重向量 $W_0 = (w_{10}, w_{20}, \dots, w_{n0})$,然后计算预测分值,并与实际值求平均绝对误差 MAE 。 MAE 代表了预测值与实际值的差值,差值越小,预测的精确度越高^[9]。然后根据一定的规则改变权重值,重新计算预测分值,并与实际值再求 MAE 。如果 MAE 小于上次的 MAE ,则最佳权重选用新的权重向量。如此迭代,直至按规则穷尽所有可能权重赋值,求出最佳权重向量。

3 系统应用分析

3.1 应用机制

美国雷克教授认为影响教育大数据挖掘应用成功的因素中,组织和人力方面的因素比技术因素影响大^[10]。因此教学大数据监测预警平台在应用中,必须要建立完善的管理应用机制,才能将大数据的优势发挥出来。教学大数据监测预警平台应用机制主要包括以下几个方面。

(1) 教学应用导向。监测预警平台必须以教学管理业务需求为导向,开发相关应用和模型,对教师、学生、课程、课堂进行相关性分析,并呈现给教学管理人员和教师,同时为他们提供预警服务,及早发现问题,调整教学行为。

(2) 大数据平台持续支撑。需要建立完善的规章制度,对教学数据采集充分授权,同时明确数据隐私保护。大数据平台应明确信息管理和使用部门,并与业务部门保持紧密联系,使得大数据平台能够对业务需求持续支撑。

(3) 预警模型协同研发。业务部门应与技术部门协同研发相关教学应用模型,积极应用模型进行实际教学管理。技术人员应重视业务部门需求,积极主动从技术上解决建模的问题,从而为模型的持续改进提供技术支持。

3.2 硬件架构

教学大数据监测预警平台由大数据处理平台和教学监测预警系统构成。大数据处理平台主要由 HADOOP 集群、数据库服务器和数据挖掘服务器构成。HADOOP 集群可以从行为日志服务器和数据库服务器中收集数据,利用 HADOOP 分布式处理程序,对非结构化和半结构化数据进行处理。其结果可以直接输出到数据挖掘服务器作进一步处理。数据挖掘服务器也可以直接从数据库服务器采集数据,直接对结构化数据进行挖掘和分析。数据挖掘服务器根据算法模型对数据进行挖掘和分析,结果输出到 WEB 服务器。WEB 服务器提供对教学监测系统的支撑。教学监测系统通过统一身份认证(IDS)实现与信息门户、教务系统、教学平台的集成,以方便管理人员和教师、学生的统一登录认证。另一方面,通过统一通信系统(UCS),实现通过一个 API 接口,利用邮件、短信、微信等多种形式进行反馈。如图 8 所示。

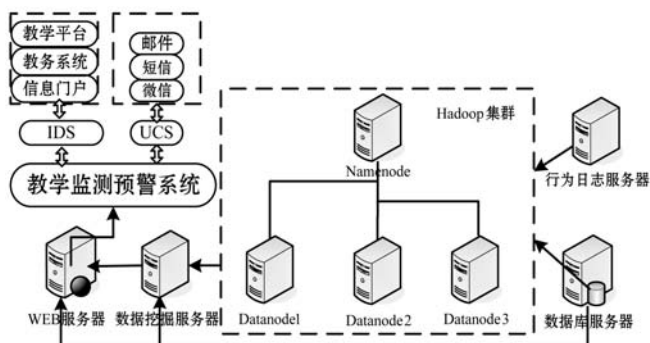


图 8 教学大数据监测预警平台硬件架构图

3.3 应用实例

本例共采集课程 4 门,每门课程约 500 多条数据。选取预警课程为运筹学(ycx),其前置课程包括高数(gs)、经济学(jjx)和管理学(glx),筛选出有效数据 404 条。利用 STATA 软件求运筹学与其前置课程的相关系数,在 $*P \leq 0.01$ 显著水平下,计算出相关系数,如图 9 所示。

	ycx	gs	jjx	glx
ycx	1.000 0			
gs	0.503 3*	1.000 0		
jjx	0.389 6*	0.602 1*	1.000 0	
glx	0.247 5*	0.311 6*	0.256 4*	1.000 0

图9 运筹学与其前置课程相关系数图

从图9可以看出,运筹学与其前置课程成绩差异显著,但运筹学与管理学相关系数 $r=0.2475 < 0.3$,不具有相关性^[11]。因此选择高数与经济学作为自变量,预测因变量运筹学的成绩。通过对2013级至2015级学生的三门成绩作多元线性回归分析,结果如图10所示。

Source	SS	df	MS	Number of obs = 404
Model	2702498.78	2	1351249.39	F(2, 402) = 7314.85
Residual	74260.2187	402	184.726912	Prob > F = 0.0000
Total	2776759	404	6873.16584	R-squared = 0.9733
				Adj R-squared = 0.9731
				Root MSE = 13.591

	ycx	Coef.	Std. Err.	t	P> t	[95% Conf. Intervals]
gs		-.3608268	.0473051	7.63	0.000	-.2678304 - .4538231
jjx		-.7006236	.0426188	16.44	0.000	-.6168401 - .7844071

图10 运筹学与其前置课程回归分析结果图

通过观察,可以看出模型整体上是非常显著的,而且解释能力也很好。自变量回归系数检验都是非常显著的,因此也可以判定两者相关系数虽高但却不存在多重共线性问题^[12]。运筹学成绩预测线性回归模型为: $P_y = 0.36 \times P_g + 0.7 \times P_j$ (P_y : 运筹学预测成绩; P_g : 高数成绩; P_j : 经济学成绩)。然后代入2016级学生成绩数据进行分析,计算平均绝对误差MAE,结果发现MAE为7,即预测成绩的误差为 ± 7 分,预测效果良好。再将该模型通过模型管理功能添加到监测预警平台,学生在选择该门课程后,就会根据该模型,产生预警信息,并自动反馈给该学生。而教学管理人员和该门课程教师也能在学生-课程预警界面,查看到相关预警信息。如此,通过对更多课程做以上建模,将不断丰富预警模型。

4 结 语

教学大数据监测预警平台以大数据平台为基础,通过对数据的分析和处理监测教学过程,并对异常信息进行预警,为教师和学生的发展性评价提供重要的技术手段。教学业务部门能够通过该平台获取教师、课程、课堂、学生的精细画像,并及早发现教学过程中的异常信息,及时反馈给相关的教师。而教师通过该平台能够在课程之初就能了解学生既往成绩和学习情况,又能根据预警信息,发现问题学生,在教学过程中给予重点关注,提高问题学生的学习效果,提高整体教学成绩。预警平台主要是针对异常信息进行提示,减轻了教学管理人员和教师的决策负担,他们不需要去

分析整体教学情况,而只要根据可视化的提示,去分析部分异常信息即可。教学大数据监测预警平台不但能提高教学效果,还能减轻管理人员和教师的教学监测压力,使得教学管理、教学过程、学习过程更为融洽、自然。

参 考 文 献

- [1] 谈松华. 关于教育评价制度改革的几点思考[J]. 中国教育学报, 2017(4): 7-11.
- [2] Daniel L S. 评估模型[M]. 苏锦丽,等译. 北京:北京大学出版社,2007.
- [3] 李景奇,韩锡斌,杨娟,等. 基于CIPP模式的网络学习跟踪与评价系统设计[J]. 电化教育研究, 2009(7): 53-57.
- [4] 李振,周东岱,刘娜,等. 教育大数据的平台构建与关键实现技术[J]. 现代教育技术, 2018(1): 100-106.
- [5] 王战军,乔伟峰,李江波. 数据密集型评估: 高等教育监测评估的内涵、方法与展望[J]. 教育研究, 2015(6): 29-37.
- [6] 陈海建,戴永辉,韩冬梅,等. 开放式教学下的学习者画像及个性化教学探讨[J]. 开放教育研究, 2017, 23(3): 105-112.
- [7] 杨现民,王榴卉,唐斯斯. 教育大数据的应用模式与政策建议[J]. 电化教育研究, 2015(9): 54-61.
- [8] 孙毅,刘仁云,王松,等. 基于多元线性回归模型的考试成绩评价与预测[J]. 吉林大学学报(信息科学版), 2013, 31(4): 404-408.
- [9] Corbett A T, Anderson J R. Knowledge tracing: Modeling the acquisition of procedural knowledge[J]. User Modeling and User-Adapted Interaction, 1994, 4(4): 253-278.
- [10] 余亮,杨秋燕,赵楠. 模型驱动的教育大数据挖掘促进教与学——访美国犹他州立大学米米·雷克教授[J]. 开放教育研究, 2018(1): 4-9.
- [11] 魏建国. 统计学[M]. 5版. 武汉:武汉理工大学出版社, 2016.
- [12] 朱顺泉. Stata 数据分析应用[M]. 北京:北京大学出版社, 2015.

(上接第58页)

- [12] Robinson I, Webber J, Eifrem E. Graph Database 图数据库[M]. 刘璐,梁越,译. 北京:人民邮电出版社, 2015: 6-33.
- [13] Sadalage PJ, Fowler M. NoSQL 精粹[M]. 爱飞翔,译. 北京:机械工业出版社, 2013: 42-50.
- [14] 康杰华,罗章璇. 基于图形数据库 Neo4j 的 RDF 数据存储研究[J]. 信息技术, 2015(6): 115-117.
- [15] 王富祥. 基于知识图谱的大数据研究可视化分析[J]. 华北理工大学学报(社会科学版), 2017, 17(1): 583-585.