

智慧校园建设研究与探索*

刘翔 行联合 李国一 许露

广州东华职业学院, 广州, 510540

摘要 在教育改革创新的推动下,各学校加快了智慧校园建设的速度,学校的各个部门在建设过程中产生了各种与教师、学生信息有关的数据,这些数据越来越多,而这些数据都与教师、学生行为相关,可以很好的反映学生的生活消费、思想情感和学习行为习惯和状况,以及教师的教学质量情况。利用校园大数据平台采用基于神经网络的数据挖掘技术对师生的日常行为数据进行分析 and 预测,为学校的管理和建设提供很好的参考价值。经过实践检验,智慧校园建设的实施方案能提高校园管理的质量和水平,为特色专业、一流专业建设提供了有力的支撑。

关键字 智慧校园, 大数据平台, 行为分析, 数据建模, 数据挖掘

Exploration and Research on Smart Campus Construction

Liu xiang

Guangzhou Donghua Vocational Institution
Institution of Information Engineering,
Guangzhou 510540, China
123171088@qq.com

Li guoyi

Guangzhou Donghua Vocational Institution
Institution of Information Engineering
Guangzhou 510540, China
21268988@163.com

Xing lianhe

Guangzhou Donghua Vocational Institution
Institution of Information Engineering,
Guangzhou 510540, China
603393376@qq.com

Xu lu

Guangzhou Donghua Vocational Institution
Institution of Information Engineering,
Guangzhou 510540, China
525916115@163.com

Abstract—Driven by education reform and innovation, schools have accelerated the construction of smart campuses. Various departments of the school have generated various data related to teacher and student information during the construction process. These data are more and more, and these data have been related to the behavior of teachers and students, and can well reflect students' living consumption, thoughts and emotions, and learning behavior habits and conditions, as well as teachers' teaching quality. The campus big data platform is used to analyze and predict the daily behavior data of teachers and students by using the data mining technology based on neural network, which provides a good reference value for the management and construction of the school. After practical tests, the implementation plan of smart campus construction can improve the quality and level of campus management, and provide strong support for the construction of characteristic majors and first-class majors.

Key words—Smart Campus, Big Data Platform, Behavior Analysis, Data Modeling, Data Mining

1 引言

随着信息技术和人类生产生活交汇融合,互联网快速普及,各行各业要与互联网结合起来才能焕发出青春活力,教育改革的发展也离不开互联网信息技术的支撑,只有这样才能促进我国教育的深化改革。当前各大校园,正在花大量的时间和资金投入如火如荼的校园信息化建设中,或者使用政府行政部门提供的软硬件,或者自己购买的软硬件,由这些软硬件构成的为各个管理部门服务的信息系统,存储了大量关于师生日常学习、生活、行为习惯的大数据。很多校

园领导的管理和决策都是依赖自己的经验或者头脑发热,很容易出现决策失误的风险,原因之一就是,领导没有大数据的思维,数据泛滥但信息匮乏,没有充分利用这些大数据的价值。如何从这些数据量大、数据类型繁多、数据价值密度相对较低的数据中挖掘出有效信息,应用于学校的管理中,为创特色、一流专业提供参考指导建议,提高学校风险预警机制,已成为各校园必须重点关注的內容。这也是智慧校园建设重要的组成部分。

大数据在教学管理中的应用主要包括数据的采集、存储、分析和数据挖掘等方面,能够为数据处理提供实时的、全样本的数据信息,从而为教学管理提供准确分析,用 AI 技术辅助学校领导和老师形成科学

*基金资助: 本文得到 2020 年广东省特色创新项目资金资助,项目编号: 2020KTSCX393 的资金赞助。

迅速的决策。校园大数据主要依托数据建模和数据挖掘两项关键技术,提供解决校园大数据技术应用问题领域的新途径,促进教学管理、教学模式、个性化教育、教学评价等方面创新^[1]。智慧校园的建设主要涉及到以下几个方面进行建设。

2 构建大数据平台

目前越来越多校园搭建智慧校园,打造一站式智慧校园应用服务,为实现学校教学、教研、科研、管理、服务等信息化打下良好基础。大学的智慧校园涉及到学校管理的方方面面,例如,智慧校园门禁,智慧校园图书馆,智慧校园人脸识别签到考勤,智慧校园公告栏,智慧校园智能消费系统,智慧饭堂等等,通过数据采集的自动化,一次采集多处共享,实现业务的数据化,构建服务多样化,实现校园“智慧”的建设。但是在智慧校园的实施过程中,出现了不少的问题,在学校内部相对独立分散的多种信息管理系统和资源,得不到统一整合,存在“信息化孤岛”问题,而且数据类型不一致,有网络日志、音频、视频、图片、地理位置信息、文本、关系型数据、非关系型数据等,这些信息资源不能有效地整合在一起,不能实现数据的分析和检索。为有效地实现数据共享、消除对数据的重复管理、数据不同步及数据类型不一致的问题,有必要构建大数据分析中心平台,把学校各个部门分别管理本部门业务的相关信息数据整合起来,实现数据采集点唯一,一次采集多处使用,所有的数据信息都可实现共享,这样就避免了多部门的重复劳动,节约了人力成本,保证了数据的标准化存储。

大数据分析中心平台的搭建目的在于帮助校园构建标准统一的数据存储和数据处理资源,围绕智慧校园开展大数据应用建设,最终形成面向管理服务的数据信息资源。校园大数据分析中心平台采用分层架构,将数据整合、数据交换与共享、大数据存储、数据分析、数据处理、大数据分析挖掘进行有效整合,贯通校园大数据管理和应用的各个环节,从而适应于多维异构场景下大数据处理要求,实现海量数据的高效应用管理要求^[2]。如图1所示,它包含以下4层:

(1) 数据采集层。使用2层采集技术,一是对传统业务系统组建的数据库、半结构化和结构化数据采集集成,基于Sqoop技术进行Hadoop系统和关系数据库之间的数据交换和抽取;二是教育公共数据的采集,包括公网的数据爬取、开放平台的数据接口、教育公共数据库的数据交换等。对采集到的数据需进行提取、转换和加载(extract-transform-load, ETL)处理,该过程包括数据抽取、转换、清洗和隐私脱敏等预处理,预处理集成后的数据进入校园大数据云存储中心。

(2) 数据存储层。提供基于NoSQL存储、云的列式存储或数据仓库存储能力,根据业务需求和快速部

署,可切换到相应的分布式存储模式,还可根据需要对以前的数据仓库进行集成。利用Hadoop、Hadoop YARN和Spark Mesos等集群资源管理框架,对各类存储数据进行多粒度信息融合,构建标准数据处理单元,为计算服务层提供统一的分析数据集。

(3) 数据服务层。针对校园大数据多源、异构、海量、价值密度低等特征,传统的计算模型难以处理。数据服务层需满足主流大数据处理框架的各种计算模型与实现方法,实现基于Storm流处理、Hadoop批处理、Spark内存处理等的高效数据处理、数据分析、数据挖掘、查询检索与机器深度学习。

(4) 数据应用层。数据应用层首先要满足智慧校园的各类大数据分析需求,包括基本的可视化与查询、预警、展示、检索等,分析结果能结合知识库应用于管理决策支持^[3]。

各层的每个子系统都包含不同的功能,不同功能模块组合成一个具有实际应用价值的大数据分析平台,实现对各种业务服务的前瞻性预测和分析,为校园管理者提供统一的决策分析支持,从而更好地突出平台本身的价值。

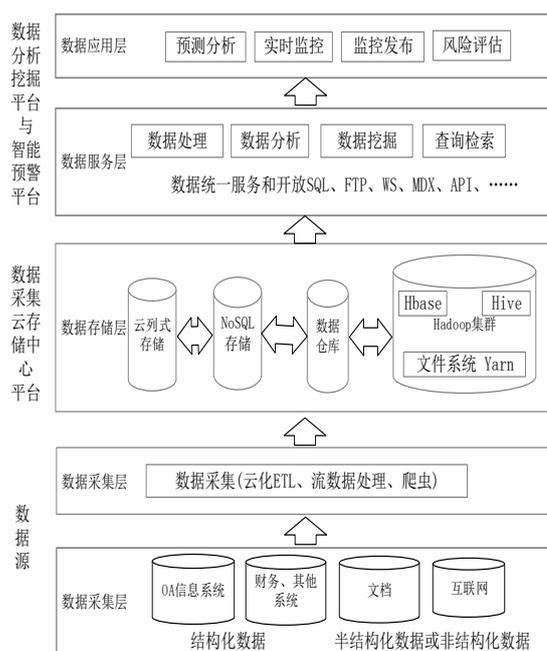


图1 校园大数据行为分析中心

3 数据建模

由图1可知,校园大数据行为分析中心的三大平台构成,即数据采集与云存储中心平台、分析与挖掘平台、监测与决策平台。数据采集与存储平台的主要任务是对学校各部门管理基础信息和师生行为相关大数据进行收集,并将采集到的数据进行结构化存储和清洗。分析与挖掘平台的主要任务是对学校采集到的

大数据进行专门的分析建模,以便后期深入对数据进行挖掘和AI分析。预警平台的主要任务是利用分析与挖掘的知识对校园或师生的异常行为进行预警等。由此可以,基于大数据的预警分析师生行为的准确性的关键在于数据挖掘的质量,数据挖掘的质量在于数据建模的精准度^[4]。

越来越多的业务需要,引进了数量急剧增长的信息管理系统,拥有了海量数据,但是分散、隔离的系统,又形成了多个孤立的数据。目前已经有了数据,如何利用数据,获得知识,于是,为了利用好数据,把不同场景的数据如何整合到同一个仓库呢?需要数据建模,也就是说,通过利用建立数据模型的手段来解决现实问题的过程。数据建模的过程是周期性循环的。数据模型通过高度抽象的数据模型,整合各个源系统的数据,最终形成统一、规范、易使用的数据仓库,进而提供包括数据集市、数据挖掘、可视化报表、实时查询等应用服务。数据模型能够促进应用需求与数据进行有效关联,形成对主要应用需求定义统一认识,可以表达和覆盖所有的应用需求。无论是操作型数据库,还是数据仓库都需要按数据模型组织数据构成,指导数据表设计。

那么,如何进行数据建模?创建数据模型的步骤可以从以下几个方面着手。

制订分析师生行为目标,准备数据。制订目标的前提是理解、了解需求,明确要解决的现实问题是什么。如:预测学校未来学生的就业率和课堂关注度。基于要解决的现实问题,理解和准备数据。明确需要哪些数据指标,即特征提取;数据指标的含义;数据的质量要求,是否存在缺失值;数据能否满足需求;数据需要加工过程,如:转换数据指标,将类别型变量转化为0-1哑变量,或将连续型数据转化为有序变量;探索数据中的规律和模式,进而形成假设。需要注意的是,准备数据的工作可能需要尝试多次。因为在复杂的大型数据中,较难发现数据中存在的模式,初步形成的假设可能会被很快推翻,需要不断试错^[5]。

根据分析学生行为的预期目标,提取特征,构建行为数据分析模型。比如学生就业率预测模型和课堂关注度分析模型。如图2所示,是右边是预测学生就业率时,所需要的数据特征,左边是预测学生关注度时需要的数据特征。就业率是学校用于评价教学质量和应届学生能力的指标,对于学校和社会间良性作用的程度具有客观的反映^[6]。建立一种基于学生在校行为的就业率预测模型,通过历史学生行为数据的比对,改进和引导学校的发展规划,促使学校的就业率稳步快速的提高。同时,通过产生可能就业困难学生名单,学校可对这些学生进行个性化指导,帮助学生就业。

建立模型。在准备好的数据基础上,建立数据模型,这种模型可能是机器学习模型,也可能不需要机器学习等高深的算法。选择什么样的模型,是根据要解决的问题目标确定。当然可以选择两个或以上的模型对比,并适当调整参数,使模型效果不断优化。

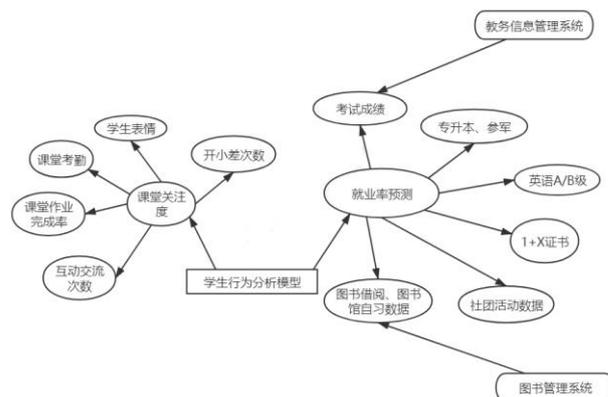


图2 学生行为分析模型数据特征指标之一

模型评估。模型效果评估有两个方面:一是模型是否解决了需要解决或潜在的问题;二是模型的精确性是否合理,误差率或者残差是否符合正态分布等。数据建模后需要评估模型的效果,一般需要将数据分为训练集和测试集,进行评估。

4 采用基于卷积神经网络 (Neural Network) 的数据挖掘技术

数据挖掘采用模式识别、机器学习 (Machine Learning) 等技术,可以从海量的、杂乱无章的数据中,发掘潜在的、有价值的知识,为智慧校园建设提供预警和决策支撑。

神经网络是实现机器学习的一种方式。机器学习可以被看成一项任务,这个任务的目标就是让机器(广义上的计算机)通过学习来获得类似人类的智能。例如人类会下围棋,AlphaGo或AlphaGo Zero就是一个掌握了围棋知识、会下围棋的计算机程序。而神经网络就是实现机器学习任务的一种方法,在机器学习领域谈论神经网络,一般是指“神经网络学习”。它是一种由许多简单单元组成的网络结构,这种网络结构类似于生物神经系统,用来模拟生物与自然环境之间的交互。神经网络是一个比较大的概念,针对语音、文本、图像等不同的学习任务,衍生出了更适用于具体学习任务的神经网络模型,如递归神经网络 (Recurrent Neural Network, RNN)、卷积神经网络 (Convolutional Neural Network, CNN) 等。除了神经网络可以实现机器学习任务外,常见的还有线性回归、决策树、支持向量机、贝叶斯分类器、强化学习、概率图模型、聚类等多种方法,可见神经网络方法只

是机器学习方法中的一种。早在上世纪 80 年代,神经网络就被提了出来,但在应用上一般采用很浅层的、很小的网络。现如今,随着数据量越来越大,计算资源越来越丰富,以及算法上的改进和优化,神经网络的层数变得越来越多,学习的效果变得也越来越好,这就是深度学习(Deep Learning),本质上就是深层的神经网络^[7]。

本文的数据挖掘技术为卷积神经网络,是深度学习(deep learning)的代表算法之一,具有较高的精确度。基于卷积神经网络的校园学生行为分析平台包括 6 个关键层次,分别是输入层、卷积层、池化层、卷积层、池化层和全连接层,如图 3 所示,为了提高精度,卷积层、池化层连续使用了两次。卷积神经网络在校园网数据挖掘过程中,每一层的功能及作用描述如下。

(1) 输入层。输入层的功能是实现校园信息的预处理,可以去掉一些噪声、偏差太大的数据等,将数据进行矩阵化,为了能够显示每一个数据包类别,对其进行归一化处理,以便卷积神经网络处理校园信息。输入层可以接收多种类型的数据,这些数据包括教务信息管理系统的信息、参加社团的数据、参加四六级考试的数据、早读数据、考研考证考公务员的数据、参加竞赛、服务外包、双创计划、大学生攀登计划等。

(2) 卷积层。卷积层通常包括两个关键操作,可以实现卷积网络的局部关联操作和窗口滑动操作。局部操作可以针对数据特征进行过滤,滑动窗口可以完成卷积神经网络特征的提取,实现卷积神经网络的特征分析,进一步改进卷积神经网络的准确度。卷积层可以采用 Sigmoid 函数,函数拥有很强的收敛性,因此可以在很短的时间内获取数据挖掘结果,避免过度拟合现象发生,可以大幅度提高校园数据分析准确度。

(3) 池化层。池化层可以压缩卷积神经网络处理的数据量,同时还可以减少网络设置的参数数量,避免卷积神经网络计算和处理时过度拟合。具体的,在卷积层处理的结果上,神经网络可以获取输入层数据包中的学生行为特征,这些特征数据采取池化操作之后就可以计算某一个局部卷积特征平均值,也可以计算最大值或最小值,利用这些值可以针对卷积层获取的特征数量进行过滤,从而可以降低分类器的计算复杂度,充分的减少过度拟合发生的概率。

(4) 全连接层。全连接层是一个分类器,其可以将神经网络经过学习和训练的结果输出到全连接层,这样就可以直接为校园大数据平台的应用层提供决策支撑。全连接层目前实现的接口非常丰富,这些接口数据包括可视化接口、图像曲线接口等,可以将其与教务系统集成在一起,实现数据的传输和共享,将其

输出到系统交互界面上,可以帮助决策者更加直观的浏览卷积神经网络的处理结果^[8]。

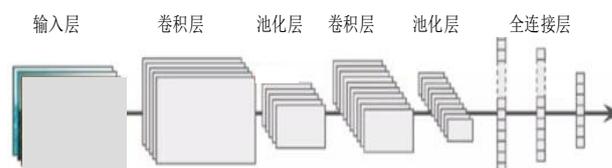


图 3 卷积神经网络处理流程

5 校园监测预警机制的梳理及归类

分析师生行为的规律与特性并建立相应的动态监测及预警系统,将学生的在校行为划分为学习行为、消费行为等几方面;分析老师的备课、教学、科研情况,对老师业绩进行评价。

(1) 学习行为分析。

收集学生在校的各种学习活动,包括学生的教务数据、档案数据、服务外包项目、竞赛、课题数据、考试数据、就业数据等进行综合性的分析,探究学生日常行为与就业情况之间的关系,为学校的学习导师和辅导员以及管理层制定规章制度提供可靠的依据。并根据学生行为建立各种预测模型,比如:挂课预测模型、留级预测模型、不能及时就业预测模型等,及时发现学生学习出现异常的情况,提前预警并及时跟进,教育改正。

(2) 生活行为的挖掘分析。

校园一卡通的消费记录,记录了学生在学校食堂、超市、奶茶店、咖啡厅的消费明细清单;校园网的上网时间,门禁记录等,反映了学生的消费水平和课外出入情况,可分析学生的业余生活,展示学生日常活动的热点区域,判断学生业余爱好和课外学习时长,以便更深入地了解学生的学习认真程度,为学生开展项目申报活动和实践提供决策参考^[9]。另外,通过消费数据能够进一步确定学生是否需要精准扶贫。

(3) 教师的业绩评估。

教师上课前的准备工作:教案,PPT,学生作业的批改;课堂上:师生互动状况,板书规范程度,讲解的熟练度,授课的逻辑、连贯性,学生的课堂关注度等;课后,学生完成作业情况,老师解答是否及时、准确;学生对教师的授课评价;学校督导对老师的评价;同部门的老师对教师授课的评价等,多方面对教师授课的业绩进行考核,此过程全部由系统自动完成,并根据对教师目前掌握的数据信息,对教师期末业绩进行有预测功能评估,给出合理化建议。当然对教师业绩的评估是全方位,多方面的。有些学校还须对教师科研方面进行考核,教师做科研的情况,申报省市省部级课题,还是横向课题,不同的课题有不同的考核

系数,课题结题情况,教师带学生参加竞赛,学院系部领导对教师的考核等。

通过对师生行为的全方位、多维度考察分析,提前预警,降低风险,提高教学质量。

6 结束语

每年众多学校要上不同的信息管理系统,学校掌握的信息越来越多,信息的更迭速度越来越快,反而有利用价值的信息变少了,原因在于大数据的利用率不高。学校把智慧校园作为其信息化建设的重点内容,校园大数据技术为其提供重要的技术支撑,可有效提升校园信息化建设水平和建设效果。

本智慧校园的建设方案提出后,在我院试行一年,学校的管理效率提高了,上了一个新台阶,学校的招生率有了大幅度提升,受到师生、企业的好评。在智慧校园的建设过程中,要充分利用大数据中的资源,深刻分析大数据发展现状和趋势,打造大数据平台,对海量、异构、多维的数据进行清洗、整合、挖掘和应用,从中提取出潜在的、有价值、极具潜在应用价值的信息,把这些信息聚集到大数据平台上,构建学生和教师行为模型,引入卷积神经网络这个数据挖掘技术,建立预警机制,能够基于有监督学习思路提高数据挖掘的精确度,为学校的教学、科研、管理、招生、就业等各项工作提供科学的数据支撑,提高了学

校的智能化和自动化水平,结合我国实际对实施国家大数据战略、加快建设数字中国作出部署要求,对于实现教育强国梦想具有重要的现实意义。

参考文献

- [1] 邓逢光,张子石.基于大数据的学生校园行为分析预警管理平台建构研究[J].中国电化教育,2017(11):60-64.
- [2] 李有增,曾浩.基于学生行为分析模型的高校智慧校园教育大数据应用研究[J].中国电化教育,2018(7):33-37.
- [3] 杜圣东,杨燕,滕飞.交通大数据:一种基于微服务的敏捷处理架构设计[J].大数据,2017,3(03):53-65.
- [4] 邓逢光,张子石.基于大数据的学生校园行为分析预警管理平台建构研究[J].中国电化教育,2017(11):60-64.
- [5] Smartbi 大数据分析平台.Smartbi 思迈特软件. [DB/OL].
<https://www.smartbi.com.cn/gn/sjmo>,2021-1-22.
- [6] 席菁.DDTR:大数据背景下的学生就业率预测模型[J].中国教育信息化,2018(21):48-50.
- [7] 机器学习,深度学习,人工智能,神经网络.CSDN 博客. [DB/OL].
<https://blog.csdn.net/Michaelwubo/article/details/79625212>,2021-1-25.
- [8] 祝晓娟.数据挖掘在医院信息系统中的应用与研究[J].网络信息工程,2020(24):80-81.
- [9] 周斌.基于大数据的大学生行为分析及预警机制研究[J].科技资讯,2020(25):33-35.