

【开放大学与远程教育】

# 面向在线教育的学习分析云平台的构建与应用

——以国家开放大学为例

魏芳芳, 魏顺平, 吴淑苹

(国家开放大学 信息化部, 北京 100039)

**摘要:** 在线教育产生的数据开放性、复杂性、多样性的特点, 使得数据汇聚与分析挖掘成为当前教育领域亟待解决的问题。学习分析云平台以大型远程教育机构——国家开放大学海量数据汇聚和处理需求为依托设计开发, 平台包含“数据源层—数据汇聚层—数据处理与学习分析层—应用服务层”等四层架构, 解决了数据多源异构、满足多用户的数据需求等技术问题, 平台部署在云端并应用于日常教学与管理工作中, 促进了国家开放大学学与教投入的持续提升。

**关键词:** 在线教育; 学习分析平台; 教育大数据; 国家开放大学; 云计算

中图分类号: TP 391 文献标识码: A 文章编号: 1008-469X(2018)05-0019-07

DOI: 10.13559/j.cnki.hbgd.2018.05.004

## 一、引言

### 1. 研究背景

中国互联网络信息中心(CNNIC)在北京发布的第40次《中国互联网络发展状况统计报告》显示,截至2017年6月,中国在线教育用户规模达1.44亿,较2016年底增加662万人。<sup>[1]</sup>面对如此庞大的在线教育规模,与传统教育方式相比,教学质量备受关注。在《国家教育事业发展规划“十三五”规划》中,“质量”一词出现了72次,并提出“制定在线教育和数字教育资源质量标准”“制定在线开放课程教学质量评价标准”(国务院,2017)<sup>[2-3]</sup>。如此规模巨大的教育群体的主要活动依赖于各类信息系统开展,必将形成在线教育大数据。如何将海量、异构的数据资源进行有效的抽取、汇聚与深入分析,对于监测在线教育过程、评价在线教育效果、保障在线教育质量均具有重要意义。

笔者所在的国家开放大学正在由一所传统远程教育机构转型升级为在线教育机构,目前拥有340万名学生,面向学生的在线学习平台每学

期产生近40亿次的请求量,连同其他招生、教务、考试、云教室等信息系统生产的数据一起,形成了名副其实的在线教育大数据。本文将以前国家开放大学在线教育大数据为基础,设计开发面向在线教育的学习分析云平台,并在国家开放大学在线教学工作中加以应用。

### 2. 研究综述

目前关于大数据分析平台的研究已经广泛应用到各个领域,杨皓等结合电子检务工程和侦查信息系统,构建了检察机关大数据分析平台,汇聚了1426791万条数据进行业务分析,提升办案水平;<sup>[4]</sup>王天军等基于Hadoop技术构建了电力行业平台,实现电力信息数据融合;<sup>[5]</sup>孙露等构建了基于沪宁高速公路的大数据平台用于ETC逃费情况分析;<sup>[6]</sup>陈建涛基于Hadoop技术设计了一款面向用户健康服务的平台,通过天气和疾病大数据进行挖掘分析。<sup>[7]</sup>可以看出,目前关于大数据分析的研究较多,但大多集中在政府、电力、医疗行业上,在教育领域的应用还比较少。

在教育领域, Jay Liebowitz 将教育大数据关键技术分为概念平台架构设计、自适应学习、

收稿日期: 2018-07-20

基金项目: 国家开放大学2018年度一般课题《面向在线教育的学习分析云平台的构建与应用》(G18F0023Y)

作者简介: 魏芳芳(1991-),女,北京人,工学硕士,实习研究员,主要从事在线教育信息整合研究。

教育数据挖掘、数据分析与可视化5个方面<sup>[8]</sup>;A Bondarev等人搭建了Hadoop架构的数据仓库,解决了教育数据聚合的问题<sup>[9]</sup>;Genlang Chen等基于Hadoop、HBase、ZooKeeper分布式计算技术搭建了适用于校园教育和在线教育的大数据平台,实现了在云计算环境下的高性能数据流转<sup>[10]</sup>;赵铮等设计了学习行为分析及推荐系统,从学习内容、路径和行为记录挖掘优秀学习者的学习路径进行个性化推荐<sup>[11]</sup>;谢延红等构建了计算机网络教学平台用于记录学习者在线教学行为,对单元测试和论坛交互行为进行了兴趣分析<sup>[12]</sup>;马晓玲等在华东师范大学使用的sakai系统上对48位学生的行为数据进行了基础统计<sup>[13]</sup>;朱珂等对sakai系统进行了二次开发,集成了数据分析工具spss对107名学生的电学课程进行了数据分析<sup>[14]</sup>;夏杨等实现了校园一卡通的大数据平台建设,实现了不在校预警,贫困生分析功能<sup>[15]</sup>;李隆帆基于大数据技术研发了MOOC教学质量评测系统,用于评估教学质量<sup>[16]</sup>;王慧在大数据、云计算基础上探讨了移动学习平台的师生交互、日志分析和用户体验方面的内容<sup>[17]</sup>;孙科针对网络教学平台SAKAI系统271名学生的行为数据进行统计分析<sup>[18]</sup>。国内在概念、学习模型、设计架构的研究较多,但对学习分析平台的研发方面还处于起步阶段<sup>[19-21]</sup>,现有的大多数平台都局限于平台建设或对指标的简单排序统计,不能描述在线学习的全过程和学习效果,数据样本也比较少。

本文构建的在线教育学习分析云平台要支持汇聚不同机构的大规模在线学生的样本数据,解决计算和存储性能不足的问题,针对关系型数据,根据不同业务部门的需求统计粗粒度数据,在细粒度数据分析上,要对学习者个体分析其异常行为数据,绘制教师和学生画像。针对非关系型数据,要进行解析转换为结构化数据,例如对发帖、课程等进行内容提取、文本挖掘、情感分析,对其分析结果进行可视化展示。

## 二、平台设计与开发

### 1. 用户需求

“互联网+”背景下的网络教育趋势是将大数据、云计算、物联网、人工智能技术的融合,让未来教育数字化和智能化。简单的统计分析不能满足教学、学习、研究和招生等工作全过程的需要,传统的各自为政的信息管理系统只能处理简单的

计算排序。针对这样的问题,我们构建了在线教育学习分析云平台,整合师生教学、学习、研究和招生数据,从用户需求的角阐述平台架构。

国家开放大学作为一个教、学、研、招一体化的组织机构,学校的在线教学工作集中在对教学现状的掌握,对教学主体的教学评价,以及适当的教学干预、数据备份与归档工作。其在线教育主体主要分为机构管理者、辅导教师、学生、研究人员、技术人员等五类代表性群体。

机构管理者的任务是监督教师和学生的行为表现,适时做出科学化的管理决策,干预教学过程,拥有最高权限。首先要了解学校整体校情,对在线学习者进行人口学统计,如性别、专业、教育背景、地域、教师和学生用户数等统计分析。其次是学情分析,对在线学习者不同分部的各项教学指标包括总行为频次、出勤率、发帖数、不同活动模块使用频次等有粗粒度统计分析及相关图表的可视化展示,对学习过程中产生的用户请求量、流量从不同的角度进行分析,评价教师教学结果和学生学习效果,绘制优秀教师和优秀学生画像,总结优秀群体行为特点,树立示范作用,绘制学习者交互图,描述学习主体学习主动性、用户之间的关系以及导出不同分部的在线教学数据。各省校管理员需要查看本机构的学习者和在线学习数据,对教师学生教学行为进行较细粒度的统计分析,绘制本机构优秀学生和教师画像,导出本机构的在线教学数据,或者对展示的数据进行二次加工,用于网络教学评估。

各省校教师关心学生学习的全过程,针对学生学习记录可以督促学生学习,进行学生特征分析、社交网络分析、内容分析和学生学习结果评价,可以查看自己担任课程的本人及学生各项统计指标结果,查看优秀学生和教师画像,查看学生出勤情况、交互情况,学生之间的交互行为分析,总结自己的工作和不足,在学生目前学习效果基础上改善教学方法,有助于教师因材施教,提高教学效果。

各省校学生可以通过平台查看自己选修课程的成绩,看到自己的行为轨迹、学习轨迹,学生之间的交互行为分析,得到自己感兴趣的相关文档、视频资源的个性化推荐,查看优秀教师和学生画像,增强自我认知意识,激发学习主动性。

研究人员根据教职工和学生的行为数据,构建用户模型,从数据中总结学习规律,量化学习

过程和结果, 评估学习效果, 预测未来学习量、辍学异常行为频率、贫困生的概率等。技术人员针对不同业务部门的需求对具体的平台开展设计与研发, 解决数据孤岛的问题, 将不同部门产生的业务数据进行汇聚与备份工作。

为完成用户需求和应用场景需求的相关任务, 我们开发了在线教育学习分析云平台, 该平台的核心技术是数据仓库的搭建, 本文建立了基于不同主题的数据仓库, 采集形成了多种异构数据源的数据, 最后通过统一的平台为用户提供基础指标的结果展示、相关业务信息统计分析与挖掘、数据可视化展示等个性化服务。

## 2. 数据仓库搭建关键技术

根据大数据系统的一般结构以及学习分析系统的一般功能要求, 针对在线教学绩效评估的主要需求, 本课题设计开发了面向在线教育的学习分析云平台, 该平台的核心技术是数据仓库的搭建。

数据仓库不同于数据库, 前者支持复杂的分析操作和决策支持, 后者用于日常事务处理。数据仓库的定义起始于W.H.Inmon发表的Building the Data Warehouse<sup>[22]</sup>, 数据仓库是一个面向主题的、整合的、相对固化稳定的、随时间变化的数据集, 支持分析与决策操作<sup>[23]</sup>。狭义上来说是指数将多源异构的数据库的集成而建立的支持多维度分析的业务化的数据集。下面按照数据流向的角度分别介绍“数据源层—数据汇聚层—数据处理与学习分析层—应用服务层”4层架构中涉及的关键技术方法。

(1) 异构数据源配置。异构数据源配置阶段主要在数据源层, 数据源主要包括来自招生系统、教务管理系统、学习平台、学习空间等平台的结构化数据以及服务器系统日志、学校外部数据等非关系型数据, 需要对这些数据来源的连接访问方式和数据结构进行系统整理。目前数据源层的各个系统存在技术架构方式多样、多版本运行、多地部署、多平台部署等问题, 给数据汇聚带来不少困难。

第一, 技术架构方式多样。由于各个平台的开发时间和开发团队不完全相同, 所采用的技术架构特别是数据库不完全相同, 存在Sql Server、Oracle、Mysql等多种数据库。这就需要数据汇聚工具能够兼容各种数据库的数据采集。

第二, 多版本运行。以教务管理系统为例,

由于历史原因, 目前存在三个版本软件并行的情况, 并且三个版本软件的数据结构还不尽相同, 这就意味着在汇聚数据时要给出三套规则。

第三, 多地部署。随着云计算的普及和在线教育业务的增长, 国家开放大学计划将所有重要业务全部部署在公有云服务上。由于公有云服务招标进展问题, 目前已购公有云资源并不足以支撑所有重要业务系统的部署, 一些业务系统如教务管理系统部署在学校本地机房, 一些业务系统如学习平台、招生平台等部署在公有云上, 两个地方并未建立专线连接, 出于网络安全考虑无法进行直接数据访问, 这也给数据汇聚服务的部署带来困难。必要时只能通过人工备份数据、文件传输, 再经由数据还原导入的方式完成数据汇聚。

第四, 多平台部署。由于国家开放大学是一个体系, 一个总部、45家分部, 每个分部都有自身诉求和个性化要求, 因此在学习平台部署上, 并不是全国使用一个平台, 而是为每个分部部署一套平台, 45个分部就有45个平台, 加上国家开放大学实验学院、行业学院等机构, 共计部署59个学习平台。当要从学习平台抓取数据时, 其实是从59个平台抓取数据。

本文的数据源层包括59个Moodle平台的以MySQL格式存储的数据库, 数据量约1.8TB, 招生、教务以Oracle存储的数据库, 数据量约2TB, 空间以SQL Server存储的数据库, 数据量约50GB。

(2) 异构数据抽取与整合。这一阶段主要任务是数据汇聚, 根据采集需求配置ETL(Extract-Transform-Load, 抽取转换与加载)采集规则、抽取计划, 数据接口层存储抽取计划内的增量数据, 关系型数据的抽取比较简单, 配置完抽取数据库链接、表、字段即可完成。非关系型数据的抽取较为复杂, 本文使用Hadoop集群分布式处理冷数据。将抽取到的历史数据进行汇聚, 使用Hive数据仓库工具建立基于学生、教师、课程、机构等不同主题的数据仓库。

数据汇聚的首要任务是配置数据源, 数据源主要是各种不同业务类型不同格式的数据库, 包括教务、招生、空间和44个Moodle平台, 数据以MySQL、SQL Server、文件、Oracle等多种格式存在, 这就需要数据汇聚工具能够兼容各种数据库的数据采集。数据源层通过Kettle为每个信息源配置一个链接, 由该链接把数据从数据源层传送

到数据处理层,此外,增量抽取和调度计划也在此阶段进行。全量数据抽取是指将一个数据表中的数据按原来的结构迁移出来,并转换成ETL工具可以识别的形式。增量数据抽取较为复杂,是自上次抽取之后将需要新增的待抽取的表中数据按照用户需求迁移出来。本文采用的增量抽取方式为时间戳方式,在源表上加上时间戳字段,每次按照用户需求抽取指定间隔时间的数据,需要首先配置源数据库、源表与目标表的映射、源字段与目标字段的映射。

在线教育学习分析云平台的数据抽取环境为开源软件Kettle,部署在阿里云上,由于学习分析系统的主要数据来源学习分析云平台部署在公有云,因此本文将Kettle服务器及数据仓库服务器部署在公有云,用于采集、汇聚和整合数据。Kettle中有两种脚本文件,transformation和job,transformation完成针对数据的基础转换,job则完成整个工作流的控制。

平台实现了59个分部的数据库抽取,定期抽取间隔某几天的数据,共12个job,94个transformation,采集数据5.6亿条。

数据仓库的建立是面向不同主题的,依据不同的分析需求来确定,保证业务和数据的紧耦合,其中,每个主题对应宏观的分析领域,对分析对象有完整一致的刻画。数据仓库的建立包括构建概念模型、逻辑模型、物理模型的三层构建。概念模型从业务角度出发,划分主题域,本文按照学生、教师、课程、机构4种类别划分。逻辑模型是具体到定义表、字段去定义实体之间的关系和属性,学生、教师、课程、机构的基本信息如名称、编码。物理模型的构建是对4类数据按照不同的维度进行小颗粒的轻度汇总,供数据展示层调用。截至2017年底,学习分析系统数据仓库已经汇聚1300余万名学生的基本信息、890万名毕业生信息、1.8亿条学生选课记录以及超过10亿条学生行为日志。

(3) 数据处理与学习分析。数据处理与学习分析主要应用统计分析、聚类、预测、关系挖掘、文本挖掘、社会网络分析等数据挖掘方法对数据进行基本处理和深度分析,并构建各类评价模型,包括个体教学/学习表现模型、课程教学实施过程评价模型、机构教学实施过程评价模型等,并计算各类统计指标。

数据处理阶段负责将抽取到的历史数据进

行汇聚,关系型数据使用MySQL数据库存放,非关系型数据使用Hive存放,Hive能够将传统SQL数据映射到Map Reduce,做成分布式集群,建立基于学生、教师、课程、机构等不同主题的数据仓库,进行个体教学或学习表现、课程教学实施过程评价、机构教学实施过程模型构建与统计指标计算,数据分析,通过各类聚类、预测、关系挖掘、文本挖掘、社会网络分析形成结果数据。

数据处理层主要负责数据清洗、转换和标准化,实现总的数据库汇聚,整合教务、招生、空间、科研和Moodle平台的关系型数据库的数据,以MySQL存放,集成日志和新闻等非结构化的冷数据以Hive方式分布式存储。Hive的正常工作依赖Driver组件、MetaStore组件和用户接口三部分,其工作方法为从Driver组件开始,Driver组件将用户请求转换为HiveQL语句,编译器根据用户查询请求在MetaStore组件中获取HIVE元数据信息,编译器通过将HiveQL语句转化为抽象语法树、查询块、逻辑查询计划、物理查询计划等转换过程最终选择最佳决策。将最佳决策解释为MapReduce作业提交到Hadoop集群上,Hadoop监控作业执行过程,然后返回作业执行结果给用户。Hive的最佳使用场景是大数据集的批处理作业。

(4) 基于业务数据应用服务。基于业务数据的应用服务将面向机构管理者、课程负责人、辅导教师、学习者、研究人员等角色,通过web页面、电子邮件、app、社交软件、短信息等多种终端,提供数据报表查看与下载、数据可视化、内容推荐、数据监测、数据共享等服务。

应用服务层是数据集市,是为了满足部门级对应用程序的分析、多维数据分析、特定指标计算的特定需求而建立,是数据仓库的子集,侧面疏解了数据仓库的访问压力。本文的数据集市从属于数据仓库,每周更新一次,并且这一阶段的数据是深度汇总数据,将多个维度集合到事实表中,便于提供查询性能。

应用服务层基于Web网页、Birt报表、Echarts图表等多种方式,绘制柱状图、饼图、雷达图、折线图、云图等可视化方法具体呈现每个指标的情况,实现校情分析、学情分析、学生画像、教师画像、舆情分析、情感分析、信息推送等多种服务。同时,还要进行数据安全和用户权限管理、多终端访问管理。

整个数据仓库构建完成后,由在线教育学校

分析云平台利用 echart、Birt、Web 应用程序等技术进行调用, 基于 PC 端、LED 展示大屏、移动端等不同设备为机构管理者、课程负责人、辅导教师、学习者用户提供即时查询、不同类型的报表、多维数据分析、多指标计算、数据监控和共享、数据挖掘应用、个性化信息推荐等服务。

### 3. 平台开发与部署

平台开发环境 Spring Tool Suite、JDK 1.8, 采用 MySQL 5.5、Hive 技术构建数据仓库存储数据, 采用 Apache+Tomcat+Spring Boot+MySQL 的技术架构进行可视化应用, 采用模块化、组件化、服务化的设计, 可跨平台部署, 多终端访问。该系统采用云计算技术 (SAAS), 通过 Hadoop、Hive 分布式集群并行计算、分布式缓存、分布式部署构架整个系统的技术框架。该平台采用 service pool 技术, 为内部、外部或第三方的其他系统提供高度灵活易用的服务接口, 为平台的外部扩展和与第三方系统集成提供了快速接入的入口。目前该平台部署在阿里云, 方便全国用户访问使用。用户可通过 LED 大屏的方式访问该平台展示业务数据和关键指标, 用 PC 端应用展示业务数据、模型设计、学生画像、质量评估, 移动端应用展示统计分析类数据。

## 三、平台应用

应用学习分析云平台开展学习分析, 可以帮助在线教育工作人员回答三类问题, 即: 谁在学、怎么学以及学的结果如何。对应的应用场景则是: 在线学习者构成分析、学习全过程分析、学习结果分析。下面就对平台的三种应用场景进行介绍。

### 1. 在线学习者构成分析

在线学习者构成分析主要是针对学校不同业务招生、毕业、考试、教务系统中学习主体的分析, 包括对教师、学生、管理者、研究人员的分析。

平台可以针对在线学习者进行基础的人口学分析, 包括年龄、性别、专业、地域等, 还可以对不同入学年份的学生人数进行基础指标统计, 从不同的维度分析在线学习者, 以表格、饼图、柱状图、折线图等多种方式进行可视化, 直观地展示学习者的特征。

### 2. 在线学习过程分析

在线学习过程分析主要对在线学习的全过程进行分析, 包括对教师学生基础数据、教师教学行为数据、学生学习行为数据, 师生交互数据、

学生之间交互数据以及学习过程中产生的其他数据等各类数据的分析。

本文主要针对核心业务数据包括用户请求量、流量, 以及学生在线学习的各项指标的分析。用户请求量与用户进入平台点击次数有关, 用户点击次数越多, 请求量越大, 并且用户点击一个页面可能产生多个请求。流量是指用户在线消耗的网络流量。

平台以时间和空间的维度绘制了按天请求量分布图, 从时间维度上解释一年中每月每天的请求量变化, 观察学生学习最频繁的时间段, 管理者可以通过这一功能从宏观上了解不同学期按天的学习表现, 为掌握学习过程起到了很好的监测作用。

平台还实现了基于请求量和流量数据的地图绘制, 颜色的深浅代表数据量的大小, 可以从空间维度上得到不同省份学习状况, 便于管理者适时干预和进行教学评价, 针对不同区域的机构提出不同的教学策略, 使得教学管理科学化。同时, 此模块平台还支持不同省份的数据钻取功能, 通过点击不同的省份实现该省份的请求量流量数据展示。

学生在线学习的各项指标包括学习人数、在线课程数、资源模块数、活动模块数、平均在线天数、人均学习行为次数、提交作业数、论坛发帖数、论坛发帖人数、论坛发帖字数、浏览资源次数等。

平台实现了不同模块之间使用频次比, 得到在 Moodle 平台 14 种模块中使用频次最高的模块, 对比分析时间维度上在一个学期的不同时期各模块包括论坛、资源、测验、论坛、作业使用的比重变化情况。还可以通过表格统计各项指标学习人数、在线课程数、资源模块数、活动模块数、平均在线天数、人均学习行为次数等指标的计算。

### 3. 学习结果分析

学习结果分析是对学习主体完成学习活动后的结果进行分析, 是在线教育各方利益相关者关心的内容, 包括师生教学现状、学习量、教师教学质量、学生学习效果的分析。根据学生学习情况, 管理者需要定期或者不定期报表产出, 进行基础的数据统计, 描述校情与学情现状; 针对不同的机构、课程和个体进行教学行为评价; 进行数据聚合, 融合历史记录和学习相关数据进行数据分析与挖掘, 例如预测未来学习量、提出相关干预策略、预测辍学风险等。

(1) 机构行为分析。在线教育学习分析云平台可以从学生活动广度、交互及时度、交互深度、交互广度、教师活动持续度、教师活动深度、教师活动广度、学生活动深度、学生活动广度等多个维度绘制不同分部的学习结果雷达图,便于不同机构之间的教学质量检测与评价。

(2) 个体行为分析。此外,平台还基于学生和教师的历史数据,建立模型,进行优秀学生 and 教师画像,将前6%的优秀学生进行群体画像,总结优秀学生的专业分布、性别、人均点击数、人均在线天数、人均行为次数、教育背景、入学年份等特征,为其他学生以及教师、管理者提供参考,树立榜样模范作用。针对不同活动、个体和机构,平台还可以绘制雷达图来展示不同维度的学习结果,

为科学评价学生教师和机构在线学习有效性提供了便利。

(3) 整体行为分析。自2015年以来该平台为国家开放大学总部、分部、学院、中心等各级机构管理者、教师、学生提供在线教学表现反馈,不断提升在线教学参与面和参与深度,在线教学表现持续改进。表1为2015年至2017年在线学生教师产生在线行为的直接效果,可以看出,自2015年以来,参与学生数、参与教师数迅速增加,学生行为数、教师行为数、学生人均行为数、教师人均行为数都不断增加,此平台在国家开放大学上的推广和应用起到积极作用,并不断提高学生和教师的学习活跃性。

表1 在线教学表现持续改进

学期	参与学生数 (环比增长率)	学生行为数 (环比增长率)	学生人均行为数 (环比增长率)	参与教师数 (环比增长率)	教师行为数 (环比增长率)	教师人均行为数 (环比增长率)
2015年秋季	498 602 (-)	38 048 082 (-)	76 (-)	8 345 (-)	3 517 306 (-)	421 (-)
2016年春季	528 094 (5.91%)	56 276 155 (47.91%)	107 (39.65%)	11 101 (33.03%)	5 722 435 (62.69%)	515 (22.30%)
2016年秋季	627 266 (18.78%)	100 776 615 (79.08%)	161 (50.76%)	14 332 (29.11%)	10 799 352 (88.72%)	754 (46.17%)
2017年春季	740 493 (18.05%)	172 356 221 (71.03%)	232 (44.86%)	18 381 (28.71%)	18 539 504 (71.67%)	1 033 (36.98%)
2017年秋季	858 692 (15.96%)	372 932 642 (116.37%)	434 (87.20%)	24 506 (33.32%)	49 583 177 (167.45%)	2 023 (95.87%)

#### 四、研究结论

在线教育学习分析云平台利用大数据技术和数据挖掘的知识,建立了国家开放大学的数据仓库,整合了多源异构的数据,平台解决了传统管理信息系统存储和计算性能不足的问题,对于在线教育的在线学习过程分析、学习结果评估、网络教学评估具有重要意义。

第一,在线教育学习分析云平台建立了数据仓库,已经汇聚1 300余万名学生的基本信息、890万名毕业生信息、1.8亿条学生选课记录以及超过10亿条学生行为日志,实现了异构教务、招生、空间系统和44个分部Moodle平台数据库的数据汇聚,解决了传统应用平台面对海量数据的性能和存储问题。

第二,完成了按照机构、课程、学生、教师的

不同角度的基础指标的计算与分析,对师生交互数据、日志数据进行分析挖掘,得到前5%的优秀学生和优秀教师的画像,为用户提供即时查询,定期更新数据服务。

第三,平台实现了学习者(人口学分析)、学习全过程(流量请求量、不同模块使用频次比)、学习结果(不同分部、个体、机构学习结果,优秀教师和优秀学生画像)的分析和可视化,使用户更直观地掌握教与学的动态,辅助用户决策。

第四,该平台涉及的数据仓库搭建、指标计算与分析、Web数据抽取与解析、数据可视化等关键技术在其他领域极具参考价值,对于网络教学评估和在线学习过程有指导作用,教师和学生可以根据自己的行为特点进行改进,以教育信息化的手段提高教学学习效果。

## 参考文献:

- [1] 中国互联网络信息中心(2017). 第40次《中国互联网络发展状况统计报告》[R]. <http://www.cnnic.net.cn/hlwfzjy/hlwzbg/hlwtjbg/201708/P020170807351923262153.pdf>, 2018-03-13.
- [2] 国务院(2017). 国务院关于印发国家教育事业发展“十三五”规划的通知(国发〔2017〕4号)[EB/OL]. [http://www.gov.cn/zhengce/content/2017-01/19/content\\_5161341.htm](http://www.gov.cn/zhengce/content/2017-01/19/content_5161341.htm), 2018-06-14.
- [3] 教育部. 教育信息化十年发展规划(2011—2020年)[EB/OL]. [http://old.moe.gov.cn/publicfiles/business/htmlfiles/moe/s3342/201203/xxgk\\_133322.html](http://old.moe.gov.cn/publicfiles/business/htmlfiles/moe/s3342/201203/xxgk_133322.html), 2018-06-14.
- [4] 杨皓, 方博. 云计算基础上的检察机关大数据分析平台构建与应用[J]. 人民检察, 2014(22):71-72.
- [5] 王天军, 沈佳, 尹蕊, 等. 基于Hadoop平台的电力行业大数据分析技术应用[J]. 电子技术与软件工程, 2016(24):177-177.
- [6] 孙露, 姜川, 柳磊, 等. 基于沪宁高速公路大数据分析平台的ETC堵逃应用探讨[J]. 中国交通信息化, 2015(s2):1-2.
- [7] 陈建涛. 面向用户健康服务的大数据平台设计及其分析处理关键技术[D]. 南京: 南京邮电大学, 2016:19-22.
- [8] Liebowitz J. Thoughts on Recent Trends and Future Research Perspectives in Big Data and Analytics in Higher Education[M]. Big Data and Learning Analytics in Higher Education. Springer International Publishing, 2017:15-20.
- [9] Bondarev A, Zakirov D, Zakirov D. Data warehouse on Hadoop platform for decision support systems in education[C]. Twelve International Conference on Electronics Computer and Computation. IEEE, 2016:1-4.
- [10] Chen G, Wen S, Yang J, et al. Building an Experimental Platform for Cloud and Big Data Education[J]. 2014:108-109.
- [11] 赵铮, 李振, 周东岱, 钟绍春. 智慧学习空间中学习行为分析及推荐系统研究[J]. 现代教育技术, 2016(1):100-106.
- [12] 谢延红, 钱爱增. 基于学习分析技术理念的计算机网络教学平台设计与实现[J]. 计算机教育, 2015(15):66-71.
- [13] 马晓玲, 邢万里, 冯翔, 吴永和. 学习分析系统构建研究[J]. 华东师范大学学报(自然科学版), 2014(2):1-19, 39.
- [14] 朱珂, 刘清堂. 基于“学习分析”技术的学习平台开发与应用研究[J]. 中国电化教育, 2013(9):127-132.
- [15] 夏杨, 王芳. 基于一卡通数据的校园大数据分析平台的构建与应用[J]. 华中师范大学学报(自科版), 2017(s1):146-151.
- [16] 李隆帆. 基于大数据技术的MOOC教学质量评测系统[D]. 大连: 大连理工大学, 2015:18-24.
- [17] 王慧. 数字校园中基于云计算的移动学习平台研究[J]. 电脑知识与技术, 2015(26):88-89.
- [18] 孙科. 基于数据挖掘的网络教学平台用户行为分析研究[D]. 重庆: 重庆理工大学, 2016:10-20.
- [19] 郑晓薇, 刘静. 学习分析模型的分类与对比研究[J]. 现代教育技术, 2016(8):35-41.
- [20] 郑新娜, 关世杰. 学习分析云平台的应用[J]. 民营科技, 2016(10):85.
- [21] 冯翔, 余明华, 马晓玲, 吴永和. 基于大数据技术的学习分析系统架构[J]. 华东师范大学学报(自然科学版), 2014(2):20-29.
- [22] Inmon W H. Building the Data Warehouse, 3rd Edition[M]. Building the Data Warehouse. John Wiley & Sons, Inc. 2002:20-30.
- [23] Bhandarkar M. MapReduce programming with apache Hadoop[C]. IEEE International Symposium on Parallel & Distributed Processing. IEEE, 2010:1.

## Construction and Application of Online Education Cloud Platform for Learning Analysis ——Taking the Open University of China as an Example

WEI Fang-fang, WEI Shun-ping, WU Shu-ping

(Information Department, the Open University of China, Beijing 100039, China)

**ABSTRACT:** The characteristics of data openness, complexity and diversity generated by online education make data aggregation, analysis and mining become an urgent issue in the field of education. Learning analysis cloud platform is designed and developed based on the demand of massive data aggregation and processing in online education data analysis of large-scale open education institutions — the Open University of China. Four levels of architecture are designed, such as data source layer, data aggregation layer, data processing and learning analysis layer, application service layer. It solves the technical problems like the diversity of data sources, multi-sources, heterogeneous, and meets the data requirements of various different users. The platform is deployed on the cloud and applied to daily teaching and management. The application of platform promotes the continuous promotion of investment and learning in the Open University of China.

**KEY WORDS:** online education; learning and analysis platform; big data of education; the Open University of China; cloud computing