

数据驱动的实验室智能管理平台^①



彭建怡, 吴湘宁, 罗勋鹤, 刘远兴

(中国地质大学(武汉), 武汉 430074)

通讯作者: 吴湘宁, E-mail: 261525449@qq.com

摘要: 为有效利用数据进行实验室管理, 研究了实验室管理模型, 提出数据驱动的实验室智能管理流程和逻辑架构. 采用标签系统将实验室数据和实验室评估 GBDT 等算法模型进行分类存储, 构建以标签系统为条件的实验室智能管理平台, 实现了批处理和场景化两种模式的实验室管理. 智能管理平台扩展性强, 能在实验室全生命周期发挥作用.

关键词: 实验室管理; 数据驱动; 标签系统; 智能管理平台; 实验室评估 GBDT 算法

引用格式: 彭建怡, 吴湘宁, 罗勋鹤, 刘远兴. 数据驱动的实验室智能管理平台. 计算机系统应用, 2020, 29(8): 67-71. <http://www.c-s-a.org.cn/1003-3254/7512.html>

Laboratory Intelligent Management Platform Based on Data-Driven

PENG Jian-Yi, WU Xiang-Ning, LUO Xun-He, LIU Yuan-Xing

(China University of Geosciences (Wuhan), Wuhan 430074, China)

Abstract: In order to make good use of laboratory data, a laboratory management model is proposed, and then a logical framework of management and data-driven intelligent process are put forward. A tagging system is used to sort management and storage data of laboratory and intelligent processing model such as the GBDT of laboratory evaluation, and an intelligent management platform based on the tagging system is constructed. Two modes of lab management, batch and scenario, are realized based on the platform. The intelligent platform has strong expansibility and can play a role in the whole life cycle of the laboratory.

Key words: laboratory management; data-driven; tagging system; intelligent management platform; GBDT of laboratory evaluation

理论教育和实践教育构成高等教育的主体. 实验室是实践教育的主要载体, 是新形势下培养高素质人才、出高水平成果、为经济建设服务的主要场所. 近年来, 高校进行了大量实验室投资建设, 配置大量实验室设备, 用以满足开展实践教学和科研的需要, 但在建设与管理过程中也存在建设论证不充分、设备管理维护更新不及时、实验室共享开放不足等问题^[1,2]. 针对这些问题, 相关学者也开展了研究. 高俊敏、姜丽伟等从制度层面提出了实验设备建设与使用方面的优化措

施, 给后续工作提供指导参考^[2,3]. RFID 及 WiFi 等技术在实验室应用, 给实验室设备信息化管理方面带来长足进步^[4-7]. 而这些传感器和新型设备投入, 使得实验室建设与管理相关数据来源也越来越多. 课程开设与变更、实验设备性能参数、价格与生命周期, 以及实验场地使用、水电使用等过程无不产生大量数据.

数据带来价值, 同样数据量增加给实验室建设管理提供了更多的方法和思路. 例如学生实验数据和设备数据能够准确衡量某一部分知识学生掌握情况, 供

① 基金项目: 中国地质大学(武汉)中央高校教改基金(本科教学工程)(2019G51)

Foundation item: The Central Universities Education Reform Fund of China University of Geosciences (Wuhan) (Undergraduate Teaching Project) (2019G51)

收稿时间: 2020-01-02; 修改时间: 2020-01-22; 采用时间: 2020-02-11; csa 在线出版时间: 2020-07-29

教学参考;设备使用数据结合维护数据,可以更好预测设备生命周期,便于进行设备更新换代;设备使用、智能插座和门禁系统结合可以有效评价实验室使用率和开放率,对实验室评估起到有效支撑作用.但目前实验室管理系统一般都是基于现有设备来实现项目管理,偏重结果,实验过程数据采集与处理较少涉及,没有有效利用起来,且系统扩展性不强.

本文从实验室管理实际出发,研究了实验室管理模型,提出数据驱动的实验室智能管理流程和逻辑架构;依据大数据思想,采用标签系统将实验室相关数据和实验室评估 GBDT 等等机器学习模型进行分类存储;构建以标签系统为条件的实验室智能管理平台,实现了实验室管理功能,应用广泛,扩展性强.

1 基于管理流程的实验室管理模型

实验室服务于教学科研工作,其管理对象包括人、事、物、信息、经费等.涉及实验室申报、建设、实验教学的全部活动,主要包括:实验室建设规划、设置,实验室管理模式与运行机制,实验室仪器设备配置与使用,实验材料与低值易耗品管理,实验室基本信息管理与档案管理,实验教学队伍建设与培训,实验教学与科研实验管理,实验室的经费使用与检查等^[8,9].由此可见,实验室管理工作较为繁杂,涉及很多方面,但这些工作又可以细分成各个具体工作,从一个具体项目角度来说,管理工作基本模型如图1所示.

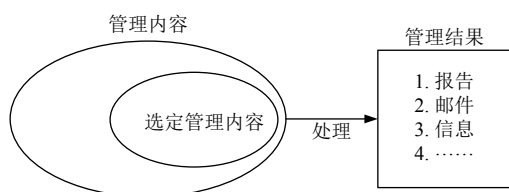


图1 实验室管理基本模型

从图1可以看出,不论管理内容有多少,对于某一具体项目管理来说,可以概括成两步:“选择”和“动作”。“选择”就是选择管理内容,“动作”就是进行合适处理后,形成报告、邮件等管理结果。“选择”管理内容操作,实际上就是限定内容满足一定要求,比如在某实验教师进行实验课程管理时,按照选择实验1成绩80分以上,并且理论课程成绩80分以上,不是补考重修学生等这些条件,筛选出指定学生,推送选修实验课程.这个过程是对具体实验对象限定,可以包含多个条件,其细分模型如图2所示.

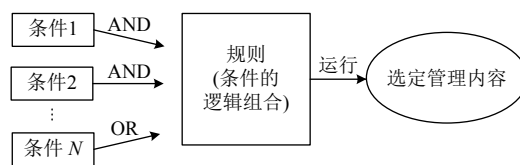


图2 选择管理内容细分模型

从图2可以看出,通过多个条件的逻辑组合,可以选择管理对象,再对该对象进行相应“动作”,其中“选择”是基础,“动作”是根据实际需要进行操作.而“选择”主要依据来源于各种数据.

2 数据驱动的实验室智能管理平台框架

根据实验室管理模型,要实现“选择”动作,需要合理利用数据来制定规则,但各个管理角色具体要求可能不一样,规则制定也不一样,对于实验教师、实验技术、系统管理人员来说,其管理流程如图3所示.

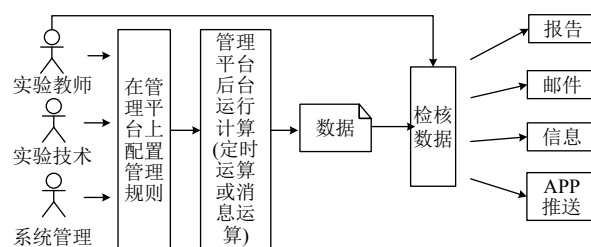


图3 数据驱动的智能管理流程

实验室管理平台分为前端和后端,前端使用Web页面,供用户操作,后端为逻辑计算,为前端展示提供逻辑和数据支持.对于数据驱动的实验室管理平台,前端供业务部门的业务人员使用,如实验教师和实验技术人员;后端的责任人为IT部门的数据工程师和系统开发工程师.根据实验室管理模型,数据驱动的实验室智能管理平台逻辑框架如图4所示.

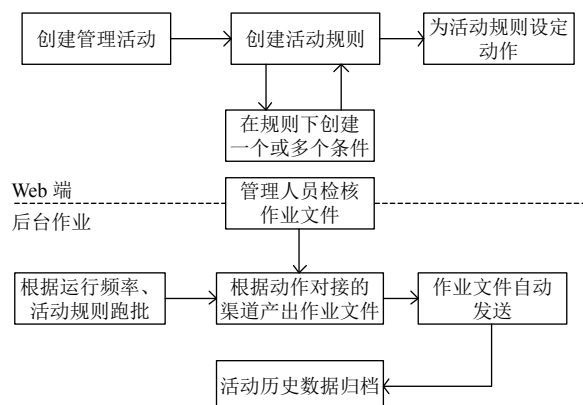


图4 数据驱动的实验室智能管理平台逻辑架构

管理人员登录管理平台,首先创建一个新的实验室管理,设置管理名称、所属责任人、时间范围和运行频率等信息,然后在此实验室管理下创建一个或多个规则,每个规则下创建不同的条件组合,最后为创建完成的规则设定一个动作.这样就完成了一个基本的数据驱动实验室管理配置.

3 基于标签系统的实验室智能管理平台实现

要实现上述管理平台实现智能管理,另外一个重要环节就是活动规则的条件设置,我们通过标签系统实现.标签,指数据标签,是对实体属性(entity)的描述,

标签的值标记了实体的一个信息.例如,对于学生,“性别”是其一个标签,“男性”是这个标签的值,它标记了该学生的性别信息.标签系统是根据一定的规则,计算和存储标签的集合,它按照既定的逻辑对标签进行管理,并根据规则进行标签值的计算和更新,通过预计算的标签解决数据访问问题,降低了数据使用门槛^[10-13].

3.1 实验室数据标签系统

标签系统是一系列标签的集合,可分为实验课程、实验设备、场地、学生、教师、管理员等主题,这些主题按数据更新方式分为基础标签,行为标签和衍生类标签,具体见表1.

表1 实验室各种数据标签

标签主题	基础标签	行为标签	衍生类标签
实验课程	名称、课程、类型(课程实验、开放实验、网络实验)、实验内容、目标、第一次开设时间、责任人...	上次开设时间、上次开放人数、本月开设课时、本月开设课时、本学期开设课时、本学期使用人数...	首次开设距天数、累计开放课时、累计开放人数、课外开放累计学时...
实验设备	名称、厂家、生产日期、价格、第一次使用时间、存放位置、责任人...	上次使用时间、上次使用时长、本月使用时长、本学期使用时长、上次维护时间、性能状况...	累计使用时间、累计使用次数、开放使用次数、外借次数...
实验场地	场地名称、大小、工位个数、第一次使用时间、存放设备数量、责任人...	上次使用时间、上次使用时长、本月使用时长、本学期使用时长、用电数据、监控门禁...	累计使用时间、累计使用次数、开放使用次数...
学生	姓名、出生日期、班级、手机号、邮箱地址、开始实验日期...	上次实验时间、上次实验内容、上次实验结果、本月实验次数、本月实验内容、本学期实验次数...	累计实验次数、累计实验学时、课内实验学时、开放实验学时...
教师	姓名、单位、人员类别(专职老师、实验兼职老师、技术人员)、手机号、邮箱、开始实验课程、首次开设课程时间...	上次实验时间、上次实验内容、上次实验结果、本月实验次数、本月实验内容、本学期实验次数...	累计实验次数、累计实验学时、课内实验学时、开放实验学时...

从表1中可以看出,标签主题分为基础标签、行为标签和衍生类标签等.基础标签的值一般固定不变,或者更新周期很长,采取增量更新的方式刷新标签值,每次仅更新值改变的标签或者插入新增加的客户端基础标签;行为标签用于描述历史行为,总是处于变化之中,采用周期性全量更新或实时(消息队列+流处理)定点更新;衍生类标签是其他标签之间的逻辑组合,它本身并不存储标签值,而是存储标签之间的计算逻辑,在被调用时才实时计算出标签值,是一种特殊的动态标签.这样实验数据均能化为标签数据,标签系统和管理系统一样也要分为前端和后端两部分,如图5所示.

标签管理员通过标签系统前端(管理)页面,配置标签与数据后台的映射关系,配置信息存储在标签映射表中.当标签用户在标签系统前端(查询)页面输入查询条件后,系统首先通过标签映射表定位标签的物理位置,然后从标签数据后台读取对应的标签值,返回页面端显示.标签映射表关联着标签系统的前端和后端,它存储了标签系统中除标签值外的所有属性,包含

对标签层次结构的全部描述信息,系统后端存储所有标签的值,通过关系型数据库表或者HBase表进行存储,标签映射表见表2,可根据存储在标签映射表中的信息序可以智能定位到对应标签的值.

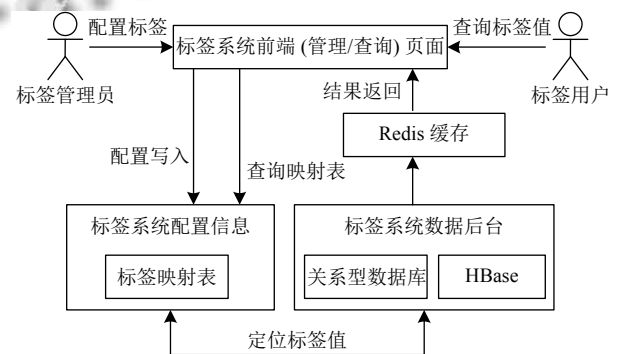


图5 标签管理示意图

3.2 标签化的实验室评估 GBDT 模型

在实验室建设论证、课程效果评估、实验室效益评估等实际管理过程中,如何有效利用数据对进行评分评价,是各方都非常关心问题.本文采用GBDT(Gradient

Boosting Decision Tree) 回归算法对前期实验数据进行建模^[14-18]. 但在 GBDT 建模阶段需要大量的计算, 因此建模过程是在后端通过离线计算完成, 并将建好的 GBDT 模型转换为标签, 供在线使用. 下面以新建实验室效益评估为例, 创建模型标签和调用流程图如图 6.

表 2 标签系统层次结构与数据后端的映射表

层次结构	关系型数据库	HBase
标签主题	Schema	表名
标签类型	表名	列族名
标签	列名	列名

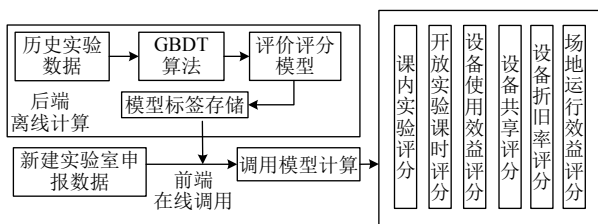


图 6 新建实验室效益评估流程图

在 GBDT 建模阶段, 在服务器后端, 根据前期存储的学生实验数据、成绩情况、设备购置价格及更新、设备使用数据、实验室用电、门禁等数据, 分别建立课内实验评分、开放实验评分、设备使用效益评分、设备共享评分、设备折旧率评分、场地运行效率评分等模型, 将这些模型以标签数据形式存储起来. 当对一个新实验室进行效益评估时, 在 Web 端调用这些模型, 输入新建实验室申报数据, 即可快速得出相应评分, 供评审专家参考.

实验室评估采用 GBDT 算法实现, 历史实验数据作为训练数据, 采用 GBDT 算法训练后, 模型函数存储为标签, 供在线调用, 其伪代码如图 7 所示.

4 基于标签系统的管理过程实例

实验室智能管理平台是通过配置管理规则来实现实验室管理. 下面以向优秀学生推送选做实验为例来说明基于标签系统的管理过程, 该项管理活动条件与标签对应见表 3.

为了配置该规则, 管理人员需要将表 3 中 3 个标签: iEX_Score1、iTH_Score、bRE_EXA 设置对应条件, 并以 AND 连接, 可以通过手工编辑逻辑关系进行调整, 图 8 展示了规则设置这个关键页面. 由此可见, 标签系统需要提供尽可能多的公用标签, 以满足尽可能多的规则 (条件) 设置要求. 一旦标签用户发现条件需要的某个标签不存在, 就需要向标签管理人员提交

新增标签需求, 当标签管理人员将新的标签增加到标签系统之后, 标签用户就可以在管理平台的上述页面看到并使用该标签.

Algorithm : GBDT of laboratory evaluation

Input :Data $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$,

Loss function $L(y, f(x))$, Gradient $\gamma, I(x)$

Output: $\hat{f}(x)$

1. Initialize $f_0(x) = \arg \min_f \sum_{i=1}^N L(y_i, f)$

2. For $m = 1$ to M :

(a) For $i = 1, 2, \dots, N$ compute

$$\gamma_{im} = - \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f=f_{m-1}}$$

(b) Fit a r_m tree, giving terminal regions $R_{jm}, j = 1, 2, \dots, J_m$

(c) For $j = 1, 2, \dots, J_m$ compute

$$\gamma_{jm} = \arg \min_{\gamma} \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma)$$

(d) Update $f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$

3. Output $\hat{f}(x) = f_M(x)$

图 7 实验室评估 GBDT 算法

表 3 条件与标签对应表

条件	对应标签名称	转换为标签后的条件
实验1成绩80分以上	实验1成绩 (iEX_Score1)	iEX_Score1>=80
理论课成绩80分以上	理论课成绩 (iTH_Score)	iTH_Score>=80
不是补考重修学生	补考(bRE_EXA)	bRE_EXA("TRUE")

规则设置页面提交时, 规则对应的逻辑关系保存至后台数据库, 其中标签以标签 ID 的形式进行保存, 通过标签 ID 可以进一步寻找标签所在的事实表, 并取得对应的标签值. 最终的逻辑关系值如果是 true, 则表明该学生满足该规则; 如果结果为 false, 则表明该客户不满足该规则, 因此不推荐该项选修实验.

界面中运行频率选项是“每天”, 指的是该规则通过页面转换成了条件之间的逻辑组合, 将条件解析成 SQL 语句, 进行后台批量处理, 结果存放在一个结果表中, 后续的管理动作阶段将根据结果表, 产出不同的格式的作业文件, 或生产报告或 Email 作业文件等过程.

除了批处理外, 还可以选择基于场景的管理, 管理活动可以基于场景的实时处理, 例如某生做完实验 1, 系统评价分数 80 分以上后, 可以直接给他进行实验推荐. 这就要将批处理改为“实时处理”, 管理过程其实是一致的, 只在技术上需要引入“消息队列”, 根据这些消息完成管理. 图 9 展示了实验提交后推荐实验效果图.

规则名称 运行频率

可用标签列表 标签主题

标签名称	标签 ID	标签值类型	所属主题	标签类型
实验1成绩	35	Integer	学生	行为标签
理论课成绩	36	Integer	学生	基础标签
补考	37	BOOL	学生	基础标签

所选条件

条件1	实验1成绩	>=	80	删除该条件
条件2	理论课成绩	>=	80	删除该条件
条件3	补考	=	TRUE	删除该条件

逻辑关系 ()

条件1 AND 条件2 AND 条件3

图8 管理平台规则设置页面

恭喜, 实验5已经完成, 您的成绩是: 95分。
根据您的学习情况, 推荐您继续完成以下实验。

<input type="button" value="选择实验1"/>	<input type="button" value="继续"/>
<input type="button" value="选择实验2"/>	<input type="button" value="继续"/>
<input type="button" value="选择实验3"/>	<input type="button" value="继续"/>

图9 基于场景的推荐实验页面

5 结论

数据带来价值, 数据带来新思路. 随着新型设备和物联网在实验室的应用, 各种实验相关数据来源越来越多, 可供管理依据也越来越多. 据此, 本文从实验室管理实际出发, 研究了实验室管理模型, 对于实际管理工作, 主要分为“选择”和“动作”两种操作, 提出数据驱动的实验室智能管理流程和逻辑架构; 依据大数据思想, 采用标签系统将实验室相关数据和实验室评估GBDT等算法模型进行分类存储; 构建以标签系统为条件的实验室智能管理平台, 实现了实验室管理功能, 应用广泛, 扩展性强. 这种基于数据驱动的实验室智能管理平台在实验室论证建设、建设流程管理、实验教学与开放使用、运行管理与维护、实验效果评估等实验室全生命周期里发挥作用, 能促进实验室发挥最大效能, 为后期建设立项提供有力支撑.

参考文献

- 陈兰, 杨经伦. 独立学院实验设备维修管理现状及发展对策. 实验室科学, 2016, 19(1): 187-189. [doi: 10.3969/j.issn.1672-4305.2016.01.056]

- 高俊敏, 袁荣焕, 刘元元. 加强高校实验仪器设备管理, 提高实验设备利用效率. 实验技术与管理, 2008, 25(9): 173-175, 182. [doi: 10.3969/j.issn.1002-4956.2008.09.051]
- 姜丽伟. 教学实验设备使用与管理的优化. 实验技术与管理, 2011, 28(10): 206-208. [doi: 10.3969/j.issn.1002-4956.2011.10.062]
- 郭鹏, 张海燕, 刘振永, 等. 基于RFID的实验设备管理系统设计. 实验技术与管理, 2015, 32(3): 255-258. [doi: 10.3969/j.issn.1002-4956.2015.03.068]
- 王法玉, 张晓洪, 陈洪涛, 等. RFID及WiFi技术在实验设备管理中的应用. 实验技术与管理, 2015, 32(5): 274-277. [doi: 10.3969/j.issn.1002-4956.2015.05.073]
- 马永军, 胡英. 基于Tek设备的智能仪器管理系统开发. 高校实验室工作研究, 2016, (4): 112-115.
- 周瑞, 李景宇, 桑楠, 等. 基于物联网的智慧实验室系统的构建. 实验科学与技术, 2017, 15(5): 146-149, 163. [doi: 10.3969/j.issn.1672-4550.2017.05.036]
- 滕利荣, 孟庆繁. 高校教学实验室管理. 北京: 科学出版社, 2008.
- Ma YC, Wang FY, Wang ZZ. Intelligent laboratory management system based on Internet of Things. Proceedings of 2017 12th International Conference for Internet Technology and Secured Transactions. Cambridge, UK. 2017. 464-467.
- 刘健, 张琨, 陈旋. 基于标签和协同过滤的个性化推荐算法. 计算机与现代化, 2016, (2): 62-65, 71. [doi: 10.3969/j.issn.1006-2475.2016.02.014]
- 李艳. 大数据环境下个性化推荐系统的研究和应用. 长春大学学报, 2019, 29(6): 34-38.
- 马勇, 鲜敏, 郑翔, 等. 基于Web日志挖掘和相关性度量的电子商务推荐系统. 计算机系统应用, 2016, 25(8): 91-95. [doi: 10.15888/j.cnki.csa.005341]
- 郭婧婧, 王颖纯, 刘燕权. 城市图书馆馆藏资源推荐系统调查分析. 图书馆学研究, 2019, (4): 76-82.
- 张潇, 韦增欣, 杨天山. GBDT组合模型在股票预测中的应用. 海南师范大学学报(自然科学版), 2018, 31(1): 73-80.
- 裴大卫. 基于深度学习的多因子股票价格预测方法研究[硕士学位论文]. 合肥: 中国科学技术大学, 2019.
- 谢天保, 赵萌. 基于网络搜索数据的游客量组合预测模型. 计算机系统应用, 2018, 27(7): 199-204. [doi: 10.15888/j.cnki.csa.006416]
- Zhang CS, Zhang Y, Shi XJ, et al. On incremental learning for gradient boosting decision trees. Neural Processing Letters, 2019, 50(1): 957-987. [doi: 10.1007/s11063-019-09999-3]
- Fu FC, Jiang JW, Shao YX, et al. An experimental evaluation of large scale GBDT systems. Proceedings of the VLDB Endowment, 2019, 12(11): 1357-1370. [doi: 10.14778/3342263.3342273]