

高等教育质量检测平台的数据分析与挖掘探析

Research on Data Analysis and Mining of Higher Education Quality Testing Platform

王妃

Fei Wang

天津外国语大学 中国·天津 300204

Tianjin Foreign Studies University, Tianjin, 300204, China

摘要: 随着社会的发展以及教育行业的进步, 高等教育的教学质量受到更多的关注, 基于此, 中国开始稳步推进高等教育质量检测平台的建设与完善, 通过数据分析与挖掘, 在实现对高等教育的系列常态检测的同时, 利于高校把握自身的发展状况, 根据数据分析与挖掘结果来实现自我的完善。论文简单论述了高等教育质量检测平台与数据挖掘, 其后就高等教育质量检测平台的数据分析与挖掘进行了探析。

Abstract: With the development of society and the progress of the education industry, the teaching quality of higher education has received more attention. Based on this, China began to steadily promote the construction and improvement of higher education quality detection platform. Through data analysis and mining, it can realize a series of normal detection of Higher Education and help universities grasp their own development status. According to the data analysis and mining results to achieve self-improvement. This paper briefly discusses the higher education quality detection platform and data mining, and then analyzes the data analysis and mining of the higher education quality detection platform.

关键词: 高等教育; 质量检测平台; 数据分析与挖掘

Keywords: higher education; quality inspection platform; data analysis and mining

基金项目: 天津市教委科研计划项目成果 (2018SK066)。

DOI: 10.12346/etr.v3i2.3486

1 引言

数据分析与挖掘在新时期获得了较快发展, 对于各个行业的提升有目共睹。高等教育质量检测平台的优化发展亦是离不开数据分析与挖掘的参与。因此需相关技术人员给予其足够的重视, 实现数据分析与挖掘在高等教育质量检测平台方面的适宜性应用。

2 高等教育质量检测平台发展概述

为充分发挥出信息技术的具体功效, 提升高等教育质量评估检测的质量与效率, 教育部与评估中心早在 2017 年即

开始构建高校教学基本状态数据库系统 1.0, 在 2009 年开始进行数据的不断采集与补充, 以此来完善与调整该数据库系统的功能与结构。在 2012 年数据库得到了全面的升级, 并建成了 2.0 系统, 该系统实现了数据关联以及逻辑校验的增加、数据真实性与有效性的增加, 审核功能等系列功能得到了进一步的加强。为在“互联网+”与大数据背景下发挥出系统更多的功效, 评估中心开始在 2016 年着手将系统平台升级为技术完备、功能最强、规模最大的数据平台。新的高等教育质量检测平台依托数据分析与挖掘技术对中国涵盖的本科院校进行了全面的数据采集, 实现了学校→院系→专业的三级数据覆盖, 国家、省市教育部门、高校可通过同一

【作者简介】王妃 (1987-), 女, 中国黑龙江哈尔滨人, 本科, 工程师, 从事数据挖掘、模式识别、舆情分析研究。

部门实现对高等教育的质量检测,且省市教育部门与高校不需进行软件开发以及系统维护。

3 数据挖掘概述

3.1 数据挖掘概念

随着社会各方面的快速发展,各种因素产生的数据信息在以几何式的速度产生与积累,这使得社会各个层面皆需新的计算工具与计算理论来协助人们从大量的数据库中选出自身所需的知识与信息。数据挖掘应运而生,其可从海量的、实际运用中的各项数据中挖掘出潜在有效的数据,且所挖掘出的数据信息可恰当解释,整体过程极为非凡。在实际生活中,各项数据信息鱼龙混杂,数量庞大,且部分数据不完整,模糊性与干扰性异常大,有着较大的随机性,因此要进行数据挖掘需消耗较多的工作量。数据挖掘作为新兴技术的一种,其与传统形式的数据库分析有着一定的不同,传统形式的数据库分析着重于数据库的查询、整理以及分析等,在进行数据挖掘时通常没有较为明确的目标,挖掘出的结果通常有较大的未知性,但是亦有实用性与有效性特征^[1]。

3.2 数据挖掘对象

在数据的不同形式存储中,数据格式以及存储方式有着较大的差异,部分不同甚至会跟随实际数据不同产生与实际应用者本身习惯的差异化,这会在一定程度上导致数据耦合性强度不足,因此需保持数据挖掘有着多种多样的对象。为利于数据挖掘的实现,通常的方式是将其应用在关系数据库中,借助数据库中各个数据间关系的研究,发现其潜在的关系。

3.3 数据挖掘功能

数据挖掘功能主要体现在描述以及预测体验两个方面。利用数据挖掘技术,不但可对现实数据实施查找、搜索等系列描述性操作,还可对一些未发生的事件借助数据与信息实施预测与检验。在预测与检测的流程中会有新模式出现,进而对决策产生较好的辅助作用。借助数据挖掘得到的系列数据信息,可广泛应用在信息查询以及信息管理,协助相关管理层做出正确的决策,进而有效调控成产过程。当前数据挖掘技术应用较为广泛的包括聚类分析、关联分析以及分类预测等。

4 高等教育质量检测平台的数据分析与挖掘

4.1 监测数据预处理

4.1.1 数据清理

①数据校验。为提升数据信息的质量,综合分析高等教育质量采集数据,依托数据项间存在的逻辑关系实施校验工作,其数据项间的校验关系,本科生与专业培养计划间主要是借助“转专业”“双学位”来实现,而专业培养计划与教职工通过“开课情况”与“教师指导毕业综合训练”联系,

校验本科生主要从其参与教师科研项目、发表学术论文、创新创业训练等方面进行数据分析与校验;针对专业培养计划,校验内容主要包括本科生交流、专业核心课程以及毕业综合训练等;针对教职工,校验内容包括教学成果、培训进修、科研奖励以及专利、论文等。校验分表内校验以及表间校验,主要包括数据取值范围、数据字典关联、唯一性与不重复校验等。对不在规则内的校验数据进行及时处理,修正其潜在的错误信息,对于难以修正的数据按照缺省处理。

②常模比较。在数据分析与校验后,数据的规范性以及质量得到了较大程度的提升,部分误报数据到了修正,但是因用户填报时对统计时间以及数据内涵理解偏差造成的数据异常,单靠数据校验难以排除。因此将高校教学质量关联的核心数据与高校对应的性质、层次与类型和中国范围内常模数据比较,对于与常模数据相差较大的数据信息进行标记,并着重分析,判断其属于异常数据后剔除并按照缺省处理^[2]。

③历史数据比较。不同学校间有着较大的差异,即便是同层次、同类型的高校,在数据方面亦会由于各自的发展重心不同而出现较大的差距,因此部分数据存在的异常很难借助常模比较来发现。因此在完成常模比较后,还需在高校内部进行历史比较,通过观察历史数据的变化趋势,标记出与实际相差较大的数据。若是有难以合理解释的差距,需按照异常数据清除,并在统计分析时按照缺省处理。

4.1.2 数据变换

在进行数据汇总时,需将不同类型的数据进行分类后汇总统计处理,建设数据立方体。将统计单位中不属于高校的各个数据项汇总,借助 AVERAGE 函数、SUM、COUNT 等方式统计高校层面的数据。

对师资入校时间以及出生日月等系列时间数据,按照统计时点,将其计算成年龄,再根据高校需求实施分类编码,比如年龄,可用 a 代替,按照 $a \leq 18$ 、 $18 < a \leq 35$ 、 $35 < a \leq 45$ 、 $45 < a \leq 55$ 、 $55 < a \leq 65$ 、 $65 < a$ 进行数据区间的划分以及异常数据的清除,其后再进行分类编码处理^[3]。

4.1.3 数据集成

合并多数据集合的数据,检测数据间存在的冲突以及重复情况,将冗余数据删除,以学校代码为统一标准进行匹配字段集成。

4.2 高等教育质量检测

进行高等教育质量检测的过程中,质量监测数据项主要包括以下多个方面:

①学校概况:学校概况包括的数据项包括本科专业情况、教职工数、管理人员数、授课教师数、教学经费、办学定位、校区数、固定资产、学生数等。

②全校整体情况:全校整体情况包括质量报告核心数据、

师资情况、教学资源极其使用情况、课程教学情况、科研仪器设备、教学经费、行政用房以及校园网等。

③各教学单位情况：包括学生以及学生管理人员情况、教育教学改革与研究情况、课程开设情况、实验技术人员、专业带头人、本科生及其教师。

④专业情况：包括学生情况、支持条件、课程体系、师资情况以及其他基本情况。

⑤全国核心数据比较：可查看中国范围内与学校同类别、同层次高校的核心数据，进行全面的对比，协助学校认清其中的差距与问题并作出对应的整改。

4.3 教育质量模型

教育过程与教育投入以及学生学习结果等产生的数据信息有着直接的关联，在全面分析高校自身教育特点以及相关数据的基础上，遵循“投入—产出”原则，综合考虑各个方面的教学评价指标，按照“教学投入→教学过程→学习结果”模式，筛选出与教学质量关联度较高的数据项为核心指标^[4]。

4.3.1 教学投入

教学投入数据项包括教学经费、固定资产、校外基地、校内实验室、图书数量、教学行政用房、教师规模、专业规模，其数据构成分别对应教学经费总额、教学科研仪器设备值、校外基地数量、校内实验室面积、纸质图书数量、行政用房与教学用房总面积、专任教师数量、专业设置数量。

4.3.2 教学过程

教学过程数据项包括创新创业培训数、培养模式创新区、课外活动项目、教授授课比例、教学班规模、实验教学学分，

其数据构成分别对应学生创新创业培训人次、学校培养模式创新试验区数量、校外活动项目参与人次、教授授课比例、课堂平均学生人数、实践学分数。

4.3.3 学习结果

学习结果数据项包括学习成效、社团活动、授予学位情况、就业情况，其数据构成分别对应学生学习成果获得数量、社团数量及其对应的学生参与人次、学位授予情况以及就业情况。

4.3.4 建设教育质量层次结构模型

以教育质量系列的数据指标，构建对应的结构模型，如图1所示。

4.4 高等教育质量检测平台架构设计

4.4.1 整合层

整合层的作用是采集高校填报数据，细化扩充评估数据以及各专业数据，再经过 ETL 数据进行清洗与萃取，最终将其集合到主数据库。

4.4.2 中间层

中间层用于数据分析，借助主数据库进行多维度数据挖掘与分析，生成各种形式的分析库，其用于数据查询、数据统计分析、数据质量监测与预警等。

4.4.3 展示层

展示层面向的是国家以及省市教育管理部门、各大高校、评估机构等，其既可作为终端发布数据信息，又可与其它平台对接，比如工程专业认证管理系统与评估管理系统等平台，以此来实现各个平台间的资源共享，如图2所示。

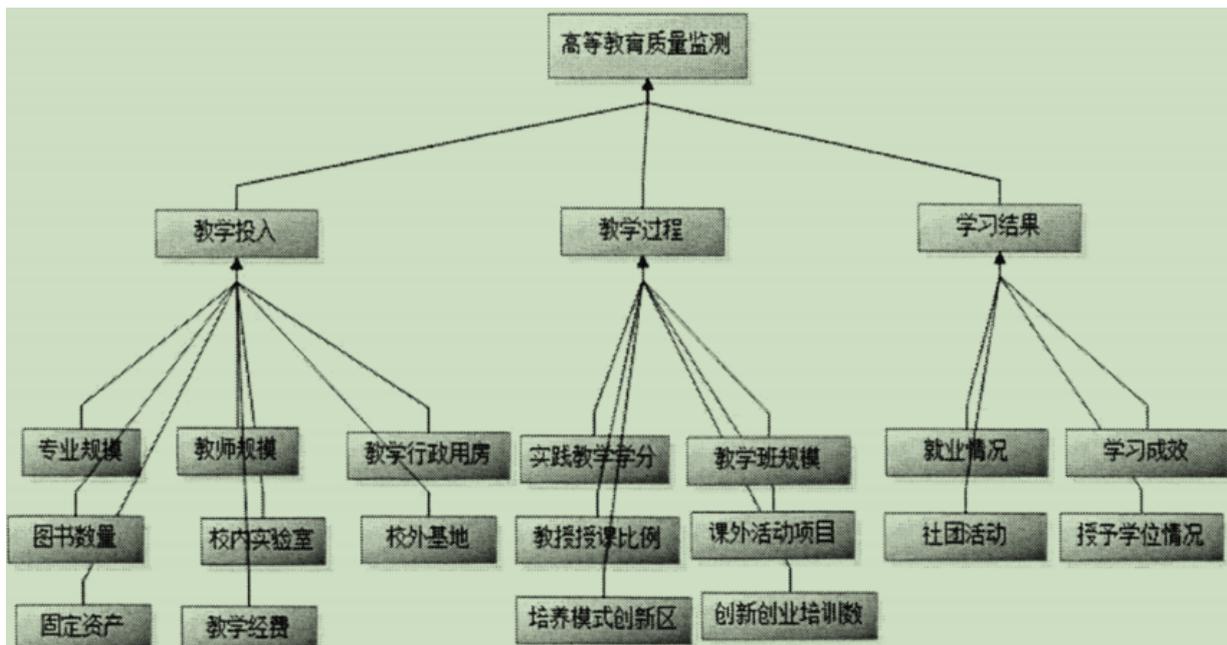


图1 教育质量层次结构模型

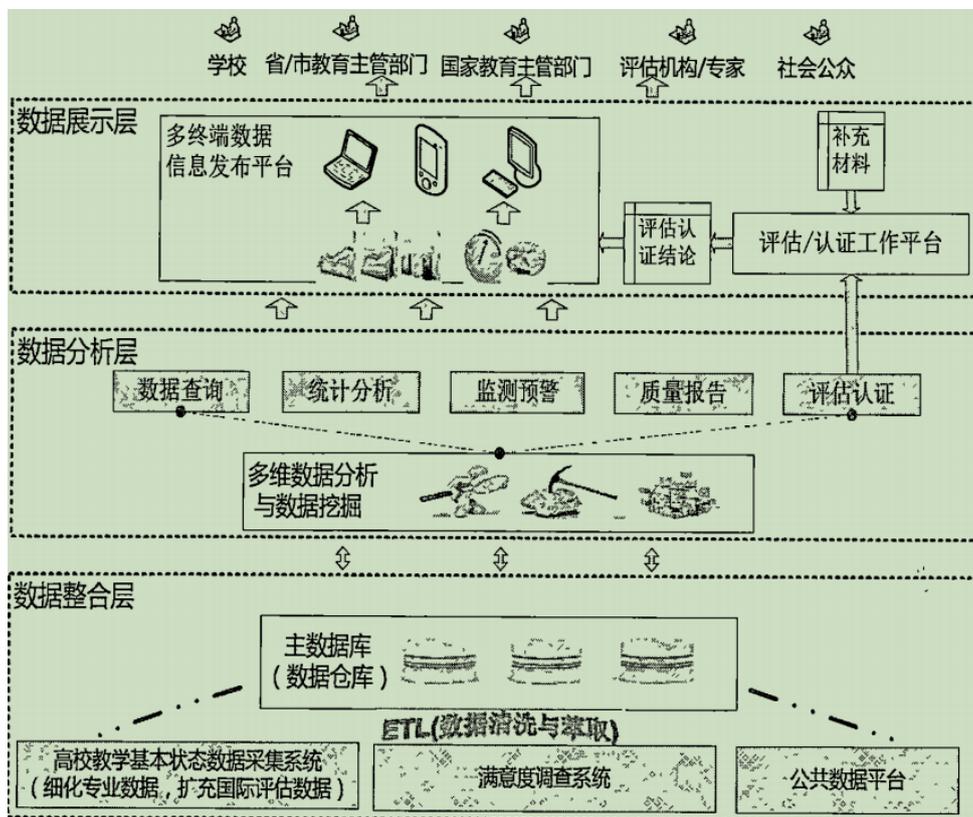


图2 高等教育质量检测平台整体架构

5 结语

综述, 实现高等教育质量检测平台的数据分析与挖掘的应用, 利于进一步调整与完善高等教育质量检测平台, 使其满足高校、教育部门以及评估中心等单位的应用要求, 且对高校的自我完善有着一定的指导意义。

参考文献

[1] 王战军, 乔刚. 以新发展理念引领高等教育质量监测数据平台建设

设[J]. 中国高教研究, 2016(7):25-30.

[2] 蔺跟荣. 高等教育质量检测平台的数据分析与挖掘研究[D]. 西安: 西北大学, 2017(19):36.

[3] 徐莹, 于耳. 构建高校大数据分析平台的探索——以上海公安高等专科学校为例[J]. 中国教育信息化, 2017, 392(5):55-59.

[4] 白云, 初庆东, 倪方映. 高等教育质量监测与评估大数据平台建设的关键问题研究[J]. 教育教学论坛, 2020(2):118-119.