Statistics and Application 统计学与应用, 2023, 12(1), 25-31
Published Online February 2023 in Hans. https://www.hanspub.org/journal/sa
https://doi.org/10.12677/sa.2023.121004

Hans汉斯

大数据技术对高等教育的影响及应用调查研究

方慧敏，李　洁，冯龙亭

阜阳师范大学数学与统计学院，安徽　阜阳

收稿日期：2023年1月9日；录用日期：2023年1月29日；发布日期：2023年2月13日

摘　要

传统的现代化理论认为现代化是社会变迁的过程，而高等教育是推动现代化社会的重要力量。如何调适自身来适应大数据时代与高等教育的现实需求，实现良好的现代转型，是目前高等教育改革与发展所面临的重要难题。本文利用主成分分析法，选择代表性变量并筛选出典型指标，通过构建多元逻辑回模型来分析当代大学生对大数据应用高等教育的满意程度，并对现状提出合理化建议。

关键词

高等教育，主成分分析，多元逻辑回归模型，满意度

Research on the Influence and Application of Big Data Technology on Higher Education

Huimin Fang, Jie Li, Longting Feng

College of Mathematics and Statistics, Fuyang Normal University, Fuyang Anhui

Received: Jan. 9th, 2023; accepted: Jan. 29th, 2023; published: Feb. 13th, 2023

Abstract

The traditional modernization theory holds that modernization is the process of social change, and higher education is an important force to promote the modernization of society. How to adapt itself to the real needs of the big data era and higher education and achieve a good modern transformation is an important problem facing the reform and development of higher education. In this

文章引用: 方慧敏, 李洁, 冯龙亭. 大数据技术对高等教育的影响及应用调查研究[J]. 统计学与应用, 2023, 12(1): 25-31. DOI: 10.12677/sa.2023.121004

paper, the principal component analysis method is used to select representative variables and select typical indicators. By constructing a multiple logistic regression model, the satisfaction of contemporary college students with the application of big data in higher education is analyzed, and rationalization suggestions are put forward for the current situation.

Keywords

Higher Education, Principal Component Analysis, Multiple Logistic Regression Model, Satisfaction

1. 引言

自第三次科技革命以来，电子计算机的诞生改变了人们的生活方式。互联网技术使文字符号转换成了数字符号，并以更加开阔的受众水平，高效的检索效率等技术特色为人们所接受和推广。2015 年，中国教育大数据研究院建立，智库的建成将实现信息大数据与教育发展的深度对接融合，对于全面深化教育改革、促进教育公平、实现教育现代化、推动教育强省和教育强国建设将发挥重要作用。2019 年，《中国教育现代化 2035》提出建设只能化校园，统筹建设一体化智能化教学，管理与服务平台。至此，教育改革在一场大数据时代背景下的已经正式开始。

2. 研究现状

虽说大数据在高等教育中的应用已开始兴起，但如何调适自身来适应大数据时代与高等教育的现实需求，开辟适合高等教育长远发展的崭新路径，实现良好的现代转型，是目前高等教育改革与发展所面临的重要难题。目前展开此类研究有左国杰(2018)结合大数据应用的课堂教学改革实验研究[1]；杨正云(2019)基于大数据技术下若干问题调查及分析[2]；陆根书(2022)大数据在高等教育领域中的应用及面临的挑战[3]；李程(2019)大数据在高等教育中的应用现状及前景[4]；李赟(2012)大数据在高校的应用研究[5]。

3. 现状分析

(一) 高等教育发展趋势

当今时代数据科学发展日新月异，大数据凭借着其大容量，高生产速度在广泛领域有着大量的应用。高等教育也随着时代的变迁而不断发展，以大数据技术作为其快速发展的前沿科技正逐渐成为一种趋势。知网一直是学术界研究领域的权威，本文通过搜索关键词为“大数据”加“高等教育”，选择 2010~2021 年的相关研究文献进行统计，并在英文数据库 ScienceDirect 以“big data”和“higher education”为主题对同期英文文献进行检索统计，可以了解国内外大数据应用高等教育的发展趋势[3]。汇总结果如图 1 所示。

由图 1 可知：中英文相关的研究文献在近十年内均有高速增长，相比较英文文献，中文文献的数量较少，增长速度比较平缓且 2011 年以后才开始出现相关文献，后面几年逐渐增长，直到 2019 年已达到 450 篇，之后趋于下降。英文文献在 2010 年已经超过 3000 篇，直到 2021 年超过了 14,000 篇的索引量。

(二) 大数据主要应用于高等教育领域

由图 2 可知，大数据应用的主要领域是教学评价、分析教育行为、质量工程、在线教育平台以及科学化管理五个领域，其中分析学生学习行为和在线教育平台的使用率相较其他三个更高。

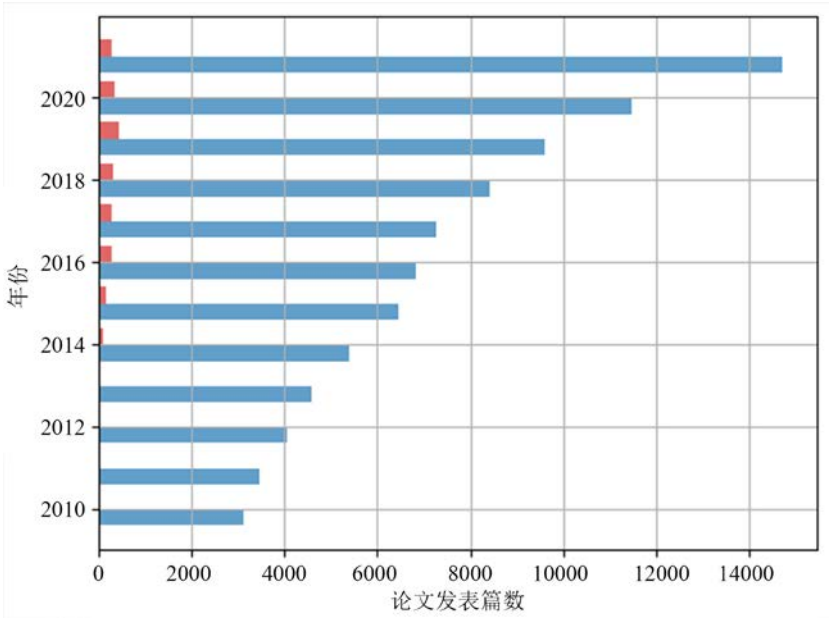

Figure 1. Big data application trend of the development of higher education (2010~2021)
图 1. 大数据应用高等教育的发展趋势(2010~2021 年)

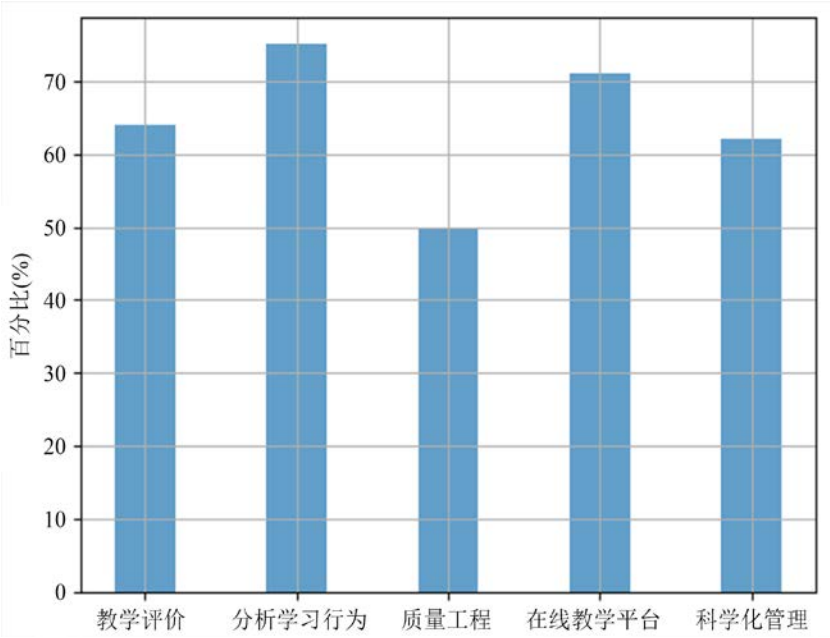

Figure 2. Big data main application fields
图 2. 大数据主要应用领域

4. 基于多元逻辑回归模型的满意度分析

(一) 各观测变量及其对应的测量问卷

依据调查问题的设计，我们认为这些问题能较好地反映学生的学习成效和满意度情况，比较适合研究，问卷中大致可以使用 5 个一级指标来进行满意度衡量。再将每个一级指标进行细分，得到 14 个二级指标，使用 123 份有效问卷作为样本，统计每一个样本的指标数据，如表 1 所示。

Table 1. Big data applications of higher education satisfaction evaluation system
表 1. 大数据应用高等教育满意度评价体系

一级指标	二级指标	具体设计
个人认知	X_1 大数据技术了解程度	学生是否需要培养大数据观念
	X_2 在生活中的运用	身边是否有从事大数据技术的人才
大数据应用	X_3 应用对象	线上学习的主要对象
	X_4 应用方法	采用何种方式进行线上学习
	X_5 数据获取方式	数据获取是否遇到障碍
(大数据)信息处理	X_6 技术支持	信息设备和教育平台的支持程度
	X_7 在线技能	大学生对是否熟练在线学习工具
	X_8 数据管理	高校是否建设数据管理中心
应用效果	X_9 教育资源管理	是否提高管理者工作效率
	X_{10} 学习途径	是否有助于合理规划学习策略
	X_{11} 课堂教学模式	与传统教学模式相比的效果
学生满意度	X_{12} 网络信息安全	使用过程中是否受到信息泄露干扰
	X_{13} 应用大学生活方面	在日常生活中是否使用相关产品
	X_{14} 线上教学方面	相比较线下学习，学习效率的影响

(二) 基于 PCA 的满意度评价指标筛选

1) 对原始数据进行标准化处理

假设样本观测数据矩阵为

$$X=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

那么可以按照如下方法对原始数据进行标准化处理：

$$x_{ij}^{*}=\frac{x_{ij}-\overline{x}_j}{\sqrt{Var\left(x_j\right)}},\ \left(i=1,2,\cdots,n;\ j=1,2,\cdots,p\right)$$

其中，$\overline{x}_j=\frac{1}{n}\sum_{i=1}^{n}x_{ij}$；$Var\left(x_j\right)=\frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij}-\overline{x}_j\right)^2;\left(j=1,2,\cdots,p\right)$

2) 计算样本相关系数矩阵

为了方便，假定原始数据标准化后仍用 X 表示，则经标准化处理后数据的相关系数为：

$$R=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix}$$

其中，$r_{ij}=\mathrm{cov}\left(x_i,x_j\right)=\frac{\sum_{k=1}^{k=n}\left(x_i-\overline{x}_i\right)\left(x_j-\overline{x}_j\right)}{n-1},n>1$。

3) 计算相关系数矩阵 R 的特征值 $\lambda_1,\lambda_2,\cdots,\lambda_p$ 和相应的特征向量

$$a_i=\left(a_{i1},a_{i1},\cdots,a_{ip}\right),i=1,2,\cdots,p$$

4) 选择重要的主成分，并写出主成分表达式

由主成分分析可以得到 p 个主成分，但是由于各个主成分的方差是递减的，包含的信息量也是递减的，所以实际分析时，一般不是选取 p 个主成分，而是根据各个主成分累计贡献率的大小选取前 k 个主成分，这里的贡献率指的是某个主成分的方差占全部方差的比重，实际也就是某个特征值占全部特征值合计的比重，即：

$$\text{贡献率}=\frac{\lambda_i}{\sum_{i=1}^{p}\lambda_i}$$

贡献率越大，说明该主成分所包含的原始变量的信息越多，主成分个数 k 的选取，主要根据主成分的累积贡献率来决定，即一般要求累计贡献率达到 85%以上，这样才能保证综合变量能包括原始变量的绝大多数信息。

利用问卷中随机抽取的 123 位大学生的相关数据，分别求出这 14 个指标的解释总方差及累计贡献率，见表 2。

Table 2. Explain the total variance and the cumulative contribution rate
表 2. 解释总方差及累积贡献率

成分	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差的%	累积%	合计	方差的%	累积%	合计	方差的%	累积%
1	2.358	18.983	18.385	2.258	18.965	18.405	2,263	15.981	15.941
2	2.236	15.678	34.174	2.186	15.147	34.192	1.896	13.356	29.368
3	1.633	11.664	45.333	1.543	11.174	45.376	1.549	10.929	40.277
4	1.464	10.332	55.866	1.454	10.582	55.828	1.167	8.267	48.598
5	1.290	9.093	64.945	1.270	9.072	64.971	1.186	7.895	56.425
6	0.954	7.082	71.977	0.964	7.031	71.923	1.091	7.439	63.828
7	0.934	6.833	78.782	0.964	6.812	78.755	1.006	7.336	71.110
8	0.876	6.172	84.897	0.826	6.114	84.838	1.063	7.304	78.437
9	0.538	4.160	89.078	0.528	4.131	89.032	1.045	7.182	85.667
10	0.530	3.727	92.813	0.530	3.783	92.824	0.959	7.136	92.874
11	0.444	2.877	95.708						
12	0.314	2.376	98.014						
13	0.276	1.969	99.958						
14	0.005	0.034	100.06						

从表 2 的结果中，前 10 个原始指标变量的累计贡献率为 92.814%。因此，前 10 个公共因子对应的代表变量作为降维后的测度指标，如表 3 所示。

Table 3. Factor analysis indicators
表 3. 因子分析选取指标

一级指标	二级指标
个人认知	X_1 大数据技术了解程度
大数据应用	X_3 应用对象
	X_4 应用方法
	X_5 数据获取方式
大数据信息处理	X_8 数据管理
应用效果	X_9 教育资源管理
	X_{11} 课堂教学模式
学生满意度	X_{12} 网络信息安全
	X_{13} 应用大学生活方面
	X_{14} 线上教学方面

(三) 基于 Logit 回归的满意度分析模型

利用上述指标，选择训练样本，采用逐步迭代的方法，得到模型总体的检验参数，如表 4 所示。

Table 4. KMO and Bartlett's test results
表 4. KMO 和 Bartlett 的检验结果

取样足够度的 Kaiser-Meyer-Olkin 度量		0.540
Bartlett 的球形度检验	近似卡方	1106.695
	df	91
	Sig.	0.000

表 4 表明，在估计模型参数时，进行到第 10 步迭代终止。−2 对数似然值(−2Loglikelihood)反映了模型中因变量不能解释的变动部分误差的显著值，Cox&SnellR 方的值在第四步分别是 0.573 和 0.803，说明模型的拟合度一般，并不是非常显著，结合表 4 综合分析，模型有一定的解释能力。

表 5 列出了变量模型的估计和测试值，除 X_8 以外，每个变量都是在 5%的显著性水平下，X_8 是数据管理，属于大数据信息处理方面，可能有很多原因出现这样显著的差异，也可能是由于样本容量波动较小导致，为了获得全面的索引信息，我们将 X_8 添加到模型中。

Table 5. Variables of the model and test value
表 5. 模型的变量及检验值

		B	S.E.	Wals	df	Sig.	Exp(B)
指标	X_3	−0.370	0.192	3.845	1	0.023	0.639
	X_5	−0.134	0.053	5.934	1	0.014	0.862
	X_8	−0.053	0.024	2.438	1	0.117	0.932
	X_9	0.431	0.145	8.825	1	0.002	1.582
	X_{11}	0.249	0.082	9.327	1	0.001	1.286
	X_{12}	−7.275	2.371	12.560	1	0.000	0.001
	X_{13}	0.002	0.001	4.456	1	0.032	1.002
	常量	1.113	1.002	1.134	1	0.025	3.046

综上所述，可得 Logit 回归模型：

$$\text{Logit}(p)=\ln\left(\frac{p}{1-p}\right)=1.114-0.038X_1-0.132X_5-0.056X_7+0.461X_9+0.259X_{10}-7.278X_{11}+0.003X_{13} \quad (1)$$

较为满意的概率为：

$$p=\frac{\exp(1.114-0.380X_1-0.132X_5-0.056X_7+0.461X_9+0.259X_{10}-7.278X_{11}+0.003X_{13})}{1+\exp(1.114-0.380X_1-0.132X_5-0.056X_7+0.461X_9+0.259X_{10}-7.278X_{11}+0.003X_{13})} \quad (2)$$

利用(2)式计算的满意度，即可衡量大学生对大数据应用高等教育的满意程度，通过与临界点(0.50)进行比较，为大数据在高等教育上的进一步应用提供依据。通过对样本进行测算得出学生总体成见很小，满意度较高，所以大数据应在高等教育方面进行深入研究，进一步完善管理，提高学生学习效率。

5. 对策与建议

1) 搭建大数据平台，优化资源配置

高校大数据应用平台通过集成高校目前的教学、科研、管理以及数字化图书馆等信息系统，对硬件设备进行统一的规划和升级，优化软硬件资源的配置，这样能为高校大数据的采集、整合和分析建立一定的硬软件基础，这也是高校大数据应用更进一步的基本前提。

2) 重视大数据人才培养，提升数据服务质量

加强对大数据人才队伍的建设，即对大数据应用人才、大数据管理人才和大数据研究人才整体队伍的建设。引进和培养大数据应用与管理人才，加强对大数据技术的应用能力，主要包括对大数据应用和管理人才的引进，对专业数据人才的培养以及对学校教师数据意识和素养提升的培训。可以利用大数据技术优化评价选项和评价细节，设置针对某一章节的教学进行评价，并且每天或者每周对评价结果进行统计分析，为教师及时调整教学方式提供参考，改善教学效果。

3) 建立大数据应用伦理规范，保证数据的有效性和稳定性

通过对大数据带来的新技术变革的“顺应”和“同化”，高等教育担负着在文化层次方面对大数据思考的责任，需从哲学、社会学、法学、伦理学等角度去规化大数据的研究和应用，以保证大数据引发社会变革的有序性和平稳性。

4) 加强政策引导，促进大数据在高等教育领域的良性发展

随着我国教育事业的发展与教育理念的进步，个性化学习将逐渐成为未来高等教育的主流方向，大数据一个用教育和主动学习将成为在高校接受高等教育时的显著特点。

基金项目

2021 年安徽省省级大学生创新创业训练计划项目：No.S202110371013。

参考文献

[1] 李方. 教育知识与能力[M]. 北京: 高等教育出版社, 2011.

[2] 杨正云, 吕亚楠. 基于大数据技术下若干问题调查及分析[J]. 法制博览, 2019(23): 70-71.

[3] 郭文革. 教育变革的动因:媒介技术影响[J]. 教育研究, 2018, 459(4): 32-39.

[4] 李程. 大数据在高等教育中的应用现状及前景[J]. 电子技术与软件工程, 2019(19): 129-130.

[5] 李赟, 胡洪都, 兰璇, 等. 大数据在高校的应用研究[J]. 现代信息科技, 2012, 3(17): 113-114+117.