

学习分析视域下的大学生 满意度教育数据挖掘及分析

舒忠梅¹, 徐晓东²

(1.中山大学 教育学院, 广东 广州 510275;

2.华南师范大学 教育信息技术学院, 广东 广州 510631)

[摘 要] 学生满意度是一个受学生和学校等多种特征影响的复杂概念,其决定因素纷繁复杂。学界对学生满意度模型的研究众多,然而从学生特征和学校特征等综合方面挖掘大学生满意度的学习分析研究较少。文章采用逐步回归和决策树分析等教育数据挖掘技术对大学生满意度数据进行分析,同时对两种方法在获得结论的有效性上进行比较,并从学习分析视角对其结果进行解释,在学生特征及学习经历的数据集中找出真正影响其满意度的维度,识别先前未知的模式,检测并分析一系列学生特征、经历和认知与满意度的关系,以期为提高高校人才培养质量提供参考。

[关键词] 学生满意度;教育数据挖掘;学习分析

[中图分类号] G434

[文献标志码] A

[作者简介] 舒忠梅(1974—),女,湖北荆门人。讲师,博士,主要从事教育数据挖掘、学习分析、高等教育管理等研究。

E-mail: issszm@mail.sysu.edu.cn。

DOI:10.13811/j.cnki.eer.2014.05.007

一、引 言

大学生满意度是大学生对于大学生活的体验和经历及对学校提供服务所产生情感的整体反应,是一个受学生和学校等各式各样特征影响的复杂概念,其决定因素纷繁复杂^[1]。最近几年,学界对学生满意度模型及其实证方面的研究较多,而较少有从学生学习经历角度来考查学生对大学学习满意度的调查。但从经验的角度分析,学生参与学校活动、经历的丰富性会对学校的满意度有一定影响,而学生会通过这些活动和经历产生各种不同的认知体验和经历及作出各种不同情感反应,涉及的数据量大、范围广,常规的统计方法难以发现其隐藏的规律,因此需要新型的数据统计和分析方法,即开展教育数据挖掘和学习分析。

另一方面,目前从学生特征和学校特征等综合方面挖掘学生满意度及其结果的学习分析研究较少,现有的研究存在样本较为局限、分析方法单一等研究限

制,也并未形成广泛认同的理论。为了深入了解学生满意度的影响因素,本文采用逐步回归和决策树分析等教育数据挖掘(Educational Data Mining, EDM)技术对学生满意程度数据进行挖掘,并从学习分析(Learning Analytics, LA)角度对其挖掘的结果进行意义再建构,旨在建立学生学习经历和学生特征与学生满意度的错综复杂关系,以期为提高高校人才培养质量提供参考依据。

二、研究分析方法

(一) 研究设计

近些年来,越来越多的国内外高校开展不同主题的学生调查,使用学生满意度等主观测量数据来辅助学校决策或反映学校办学绩效。中山大学学生学习状况调查问卷,参考 Astin^[2]与 Pace^[3]的学生学习评估模型,调查内容不仅关注学生的相关经历,例如学习投入和努力的程度、学生与教师的交往等,还包括学生

基金项目:全国教育科学规划教育部重点课题“高等教育学生学习经历分层与流动研究”(课题编号:DIA130303);国家自然科学基金项目“ITS 中基于有向超图的个性化的学习过程及其支持资源的优化”(课题编号:61202345)

的情感性学习成果或通识教育的学习成果,以及学生满意度等。

本文从学生学习经历角度,在“生源—学习—成果”的逻辑框架中,考察分析学校因素和学生因素对于学生满意度的影响机制。调查将学生学习经历分解为生源情况、学校学习资源供给、学生与学校的融合、学生学习投入、学生成果及学生满意度等六大维度。各维度下题目的内部一致性系数均达到 0.9 以上,具有较高的信度^[4]。研究数据来源于 2012 年实施的“中山大学学生学习状况调查”项目的在线调查,覆盖全校 36 个学院(系),调查总体约为 3.3 万名本科生。学生在无压力情况下作答,共回收问卷 7051 份,回收率约为 21.2%,与国际上同类的问卷回收率相比,该回收率为非常满意。

研究聚焦于本科样本,全部回收的问卷根据答题时长、问卷质量标准等原则,筛选出有效问卷 6673 份,有效率为 94.6%,涉及性别、年级、所在校区等九个人口学变量。调查收集了一系列学生特征、学习经历和规划,对校园环境、学风、服务和设施的满意度,对于成长的认知和选择大学的原因等 227 个变量数据。例如,33 个变量测量学生对校园服务、学校设施和校园文化的满意度,如“图书馆工作人员的服务质量”和“奖、勤、补、助、贷等学生资助工作与服务”等。

大学的使命是为学生提供教育,保证消费者对教育的满意度。该调查提供了在校经历满意度、能力培养满意度和综合满意度等三方面的指标来测量学生满意度,有助于在校园生活中找出影响满意度的具体方面;同时,强调了学生作为消费者的教育成果。其中,学生在校经历满意度指标通过六个方面的具体项目来反映:课程安排和教学质量、在校期间的个人生活经历、在校期间的学术经历、在校期间课外社团活动经历、在校期间参与的党团活动经历、在校期间参加的社会服务活动等;学生能力培养满意度指标主要从三方面的能力培养进行考察:获取知识能力、创新能力和社会能力;学生综合满意度指标包含五方面的项目:为学校感到骄傲、对学校各方面都感到满意、对学校有很强的归属感、会把本校推荐给他人报考、如果可以重新选择报考大学是否还会选择该校等。三个满意度指标测度在显著性水平 Sig. 为 0.01 时都是中度相关的,见表 1。

为了弄清学生对这些满意度问题有不同回应的原因为,需要了解学生的特定学习经历。对学生的学习经

历的分析有助于学校决策者制定提高学生满意度策略的优先顺序。

表 1 学生满意度的相关性

| | 综合满意度 | 能力培养满意度 |
|---------|---------|---------|
| 在校经历满意度 | 0.591** | 0.495** |
| 综合满意度 | | 0.573** |

注:**表示在 0.01 水平上显著相关

(二) 分析方法

教育数据挖掘方法研究如何有效地利用教育系统中的数据,更好地了解学生及其学习环境,常用于分析网站登录数据、学习管理系统或适应性学习系统中的学生学习数据。而在确定学生满意度模式中,教育数据挖掘也同样展示出了其潜在价值,例如,在数据集中找出学生特征及学习经历中真正影响其满意度的维度,识别先前未知的模式,如学生意见数据中的满意度模式,为教育数据挖掘在高等院校研究中的应用提供了机会。

学习分析技术围绕与学习者学习信息相关的数据,运用不同的分析方法和数据模型来解释这些数据,根据解释的结果来探究学习者的学习过程与情境,发现学习规律;或者根据数据阐释学习者的学习表现,为其提供相应反馈,从而促进更加有效的学习^[5]。

教育数据挖掘关注如何从学习的数据集中提取有用信息的技术,而学习分析侧重于如何优化学习及其学习环境的教育问题。Siemens 等学者指出,教育数据挖掘和学习分析可以联合起来共同提供更好的服务^[6]。因此,为确定学生满意度多变量数据集中显著的模式,本文以满意度理论为基础,借鉴国内外高校学生满意度模型研究成果,采用决策树和逐步回归分析等教育数据挖掘技术,检测影响大学生满意度的显著变量,并从学习分析视角对其结果进行解释,由此归纳出一系列广泛的学生特征、经历和认知与情感反应即满意度的关系。

三、学生满意度的多元回归分析

学术界常用回归方法分析现象之间的相关性,确定因果联系,并用数学模型来表现具体关系。为找出影响学生满意度的重要因素,本文采用了逐步回归方法。对于“中山大学学生学习状况调查”中的 227 项数据进行相似项合并,用向前和向后逐步回归确定与满意度结果强相关的项目,全面审查残差图,最后确定 20 个独立变量出现在多元回归模型中,见表 2。修正

来源于美国教育数据挖掘官方网站: <http://www.educationaldatamining.org>

后的校正判定系数 R^2 的范围在 0.972 到 0.982 之间,说明这些满意度模型的解释能力较强。三方面的学生满意度测量分别有不同的预测变量(显著性水平 $Sig. < 0.05$),每个模型代入相应的预测变量运行得到 β 值。

表2 学生满意度的预测*

| | 在校经历 满意度 | 综合满 意度 | 能力培养 满意度 |
|-------------------|-------------|-----------|-------------|
| 学生生源情况 | | | |
| 性别 | 0.003 | 0.025 | |
| 年级 | | 0.026 | 0.022 |
| 父亲最高学历 | 0.033 | | |
| 学校学习资源提供 | | | |
| 教师授课 | 0.072 | 0.032 | 0.064 |
| 图书馆资源 | 0.032 | 0.030 | 0.053 |
| 学生学习资源 | 0.105 | 0.244 | 0.208 |
| 学术支持 | 0.035 | | |
| 专业学习经历 | 0.185 | 0.068 | 0.089 |
| 学生学习投入 | | | |
| 自主学习 | -0.081 | -0.041 | -0.042 |
| 活动参与 | 0.158 | 0.066 | 0.048 |
| 课外阅读 | -0.018 | | |
| 朋辈交流 | 0.090 | 0.060 | |
| 讨论关注的内容 | -0.044 | 0.025 | 0.041 |
| 师生交流 | 0.075 | | |
| 社会活动时间分配 | 0.030 | | |
| 校园文化 | | | |
| 平等文化 | 0.054 | 0.129 | |
| 多元能力的培养氛围 | 0.205 | 0.216 | 0.380 |
| 学生学习成果 | | | |
| 大学生涯发展打算 | 0.050 | | 0.022 |
| 个人目标 | | 0.098 | |
| 自我能力增长 | 0.117 | 0.073 | 0.134 |
| 变量解释的百分比(R^2) | 97.2% | 98.2% | 97.4% |

注:*表示所列数字是标准化的 β 系数,对于所有独立变量, $p < 0.05$

从学习分析的角度,结合表2中的数据可以得出如下结论。

(1)学生学习资源和多元能力的培养氛围等都是影响在校经历满意度、能力培养满意度与综合满意度的重要预测变量。其中,多元能力的培养氛围指标主要包括七方面的项目:学校注重对学生批判性思维能力的培养、对学生研究能力的培养并提供参与学术研

究的机会和规范的科研环境、对学生创造能力的培养、对学生人文艺术素养的熏陶、对职业竞争力的培养、对学生信息技术的运用能力的培养、对学生的道德修养的培养、对学生的社会服务意识的培养等。

(2)对于在校经历满意度测量,多元能力的培养氛围是最重要的预测变量,专业学习经历和活动参与分别为第二、第三重要的预测变量。

(3)对于综合满意度测量,学生学习资源为最重要的预测变量,多元能力的培养氛围和平等文化分别为第二、第三重要的预测变量,表明校园文化对于学生的综合满意度影响较大。

(4)而对于能力培养满意度测量,多元能力的培养氛围为最重要的预测变量,学生学习资源和自我能力增长分别为第二、第三重要的预测变量。

综合来看,一方面,在校经历满意度、能力培养满意度与综合满意度的影响因素互不相同,通过不同的模型,分别从不同角度反映了学生的学习经历和认知情况;另一方面,学校学习资源提供和校园文化这两大维度对三方面的学生满意度均有比较重要的影响。因此,为了促进更加有效的学习,高校应该从学习资源配备及校园文化氛围的营造方面重点投入和建设,以提高学校人才培养质量,全面提升学生的满意度。

四、采用CHAID方法的决策树分析

上述多元回归分析筛选出解释因变量差异的预测变量,确定了与学生满意度相关的影响变量。决策树分析则从另一个视角,采用CHAID等算法通过 χ^2 统计检验测量因变量与自变量间的关联,发现差异,找出最能区分学生是否满意大学经历的具体方面。

CHAID程序从找到那些与因变量或目标变量有显著关系的自变量开始,然后对连续变量估计分组类别或区间间隔以选择最显著的组合。与目标变量相关性最强的自变量为树的第一个分支,而相对结果变量显著不同的每一类作为一个叶子节点。重复此过程,以找到每个叶子节点与结果变量最显著的预测量,直到没有显著预测量存在。

分析使用决策树中的“彻底CHAID”方法,包含了数据集中的每个数据点,且每个数据点只属于一个叶子节点,所以数据的分支或叶子节点是可穷尽的。然而,随着树规模的增长,最新的树层次将削减每个叶子节点的观测量。决策树程序允许分析者设定每个叶子节点的最小观测次数以避免对具有过少观测量的叶子进行 χ^2 检验,同时允许限制父节点的最小观

测量,以防止从具有较少观测量的叶子节点产生分支。鉴于研究的样本规模较大,因而限制父节点的最小观测量为 100 或叶子节点的最小观测量为 50,以防止产生新的分支,决策树的层次设置为 3~4 层。限于篇幅,下面仅对具有代表性的学生能力培养满意度进行决策树分析。

最能区分学生是否对能力培养感到满意的是多元能力的培养氛围,表 3 展示了决策树第一层叶子节点的频率分布。多元能力的培养氛围与学生能力培养满意度之间的相关性 χ^2 统计为 1852.943,比其他任何变量都高。与能力培养满意度相关的多元能力培养氛围的总体情况是:25.2%的调查对象认为其能力培养是满意的,但学生对能力培养这个问题的回应分布却因多元能力的培养氛围不同而异。例如,样本中认为多元能力的培养氛围优秀的学生有 21.1%,其中 47.7%的称对能力培养满意;而评价多元能力的培养氛围差的学生中只有 6.5%对能力培养满意。

尽管决策树分析不能确定因果关系,但值得注意的是,这里确定多元能力的培养氛围为学生能力培养满意度的一个主要决定因素与回归结果是一致的。下面层次的决策树分支进一步展示了聚类分析的价值,这种聚类分析为不同类型的学生确定不同的满意度预测变量。不同类型的学生,如认为多元能力的培养氛围分别为优秀和差的学生,受不同方面的大学经历影响。Terenzini 和 Pascarella 在调查学生特征、社会及学术整合措施的相互作用时,强调了影响因素的多样性,并呼吁将注意力集中到研究大学经历的因素变化方面^[7],但这种差异尚未得到广泛研究。

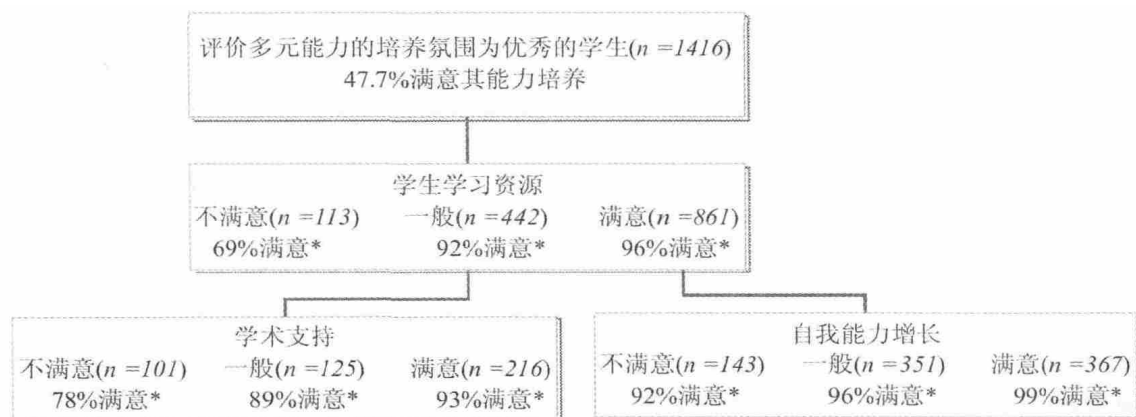
通过总结表 3 中频率分布类型的其他独立变量,图 1 确定出可以区分 1416 个称多元能力的培养氛围优秀学生的满意度变量。每个单元中显示的百分比是该组评价多元能力的培养氛围为优秀的学生的百分

比。例如,861 个称多元能力的培养氛围优秀的学生同样对学习资源满意,96%的学生满意其能力培养,而不满意多元能力的培养氛围的 113 个学生中只有 69%的满意其能力培养。继续沿着决策树分支往下查看,满意学生能力培养、认为学习资源优秀并满意自我能力增长的学生 99%认为其能力培养是满意的。考虑到分析中的限制——父节点最小观测数 100、子节点最小观测数 50、决策树的层次为 3~4 层,在不太满意学习资源的学生中没有显著的关系,所以这些叶子没有产生新分支。

表 3 学生能力培养满意度与多元能力的培养氛围

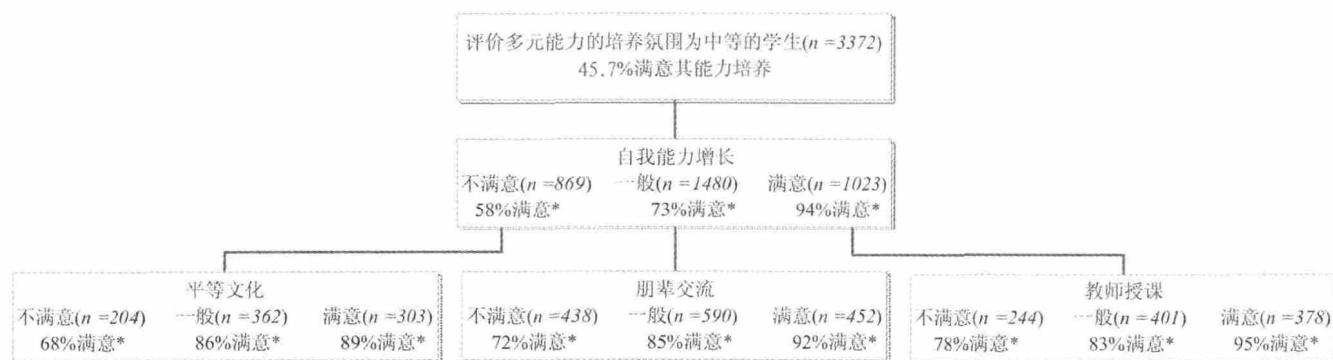
| | 多元能力的培养氛围($\chi^2=1852.943$) | | | |
|-------------|--------------------------------|-------|-------|-------|
| | 所有学生 | 优秀 | 中等 | 差 |
| 能力培养满意度的百分比 | | | | |
| 满意 | 0.252 | 0.477 | 0.457 | 0.065 |
| 一般 | 0.451 | 0.162 | 0.621 | 0.217 |
| 不满意 | 0.297 | 0.064 | 0.37 | 0.567 |
| 样本的百分比 n | 1.00 | 1.00 | 1.00 | 1.00 |
| | 100% | 0.212 | 0.505 | 0.282 |
| | 6673 | 1416 | 3372 | 1885 |

决策树其他分支如图 2、图 3 所示,分别给出了评价多元能力的培养氛围为中等的学生(样本的 50.5%)的分支,及评价多元能力的培养氛围为差的学生(样本的 28.2%)的分支。与图 1 类似,每个单元中显示的百分比是该组评价多元能力的培养氛围为优秀的学生的百分比。除了学习相关因素之外,师生交流、朋辈交流等方面的变量也是影响学生对能力培养满意度的预测量。图 2 表明,学校平等文化和朋辈交流等认知,与评价多元能力的培养氛围为中等的学生的相对满意度有关。图 3 表明,师生交流和交往圈子氛围等对于评价多元能力的培养氛围为差的学生是



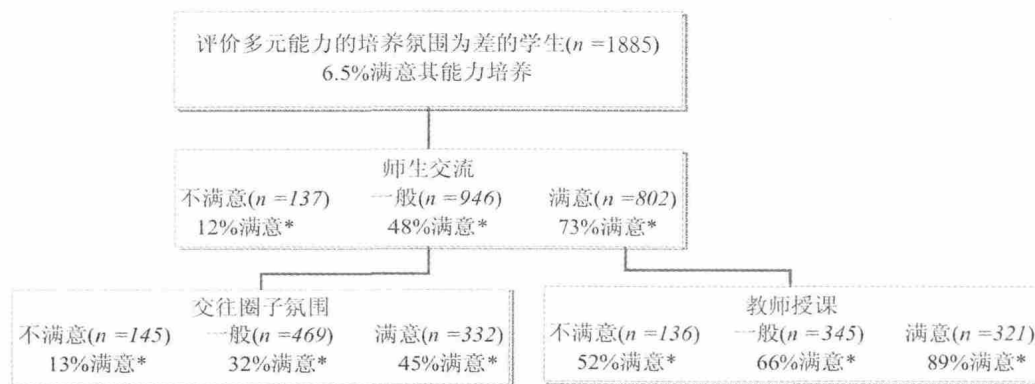
注:*表示评价能力培养满意度为“满意”的学生的百分比

图 1 学生能力培养满意度:评价多元能力的培养氛围为优秀的学生



注:*表示评价能力培养满意度为“满意”的学生百分比

图2 学生能力培养满意度:评价多元能力的培养氛围为中等的学生



注:*表示评价能力培养满意度为“满意”的学生百分比

图3 学生能力培养满意度:评价多元能力的培养氛围为差的学生

重要的。

Tinto 认为学校系统内部具有两个子系统:学术系统和社交系统。学术系统代表学生个人的课业表现、智力发展、学业成就等综合表现,社交系统代表学生在校内的同伴关系、师生关系、社交行为等综合表现;学生进入学校后需要融入学术系统和社交系统^[8]。图1~3展示了不同的大学经历如何影响对多元能力的培养氛围评价不同的学生的满意度,诸如朋辈交流、师生交流和交往圈子氛围等社交活动对较少参加学术活动的学生来说是重要的。尽管作为重要元素出现在决策树分析中的大多数变量在回归分析中也被确定为重要的,但决策树通过找出具体因素影响来澄清其显著性。上述决策树分析的学生满意度状况与Tinto对留级的建议相吻合:社交系统的融入可以弥补弱弱的学术参与。

五、学习分析及启示

本文采用决策树分析和逐步回归等教育数据挖掘技术来确定学生特征和学习经历与满意度的关系,识别多变量数据集中的显著模式,而不是检测关于预选变量影响效果的特定假设。对更广泛的学生满意度影响因素使用卡方自动交互检测(CHAD)算法,探索

分析识别学生特征、学习经历和认知与学生满意度的关系。

从学习分析的视域来看,利用多元回归分析方法,识别出影响学生满意度的重要影响因素;而决策树发现更多的学生满意度相关数据。上述决策树分析结果与回归分析结果既有类似又有不同。多元能力的培养氛围、学生学习资源、自我能力增长等作为显著影响的变量出现在最终的回归分析模型中,同时也在决策树分析中作为与学生能力培养满意度有重要关系的具体经历而出现。然而,除了这些变量,决策树分析还强调不同的二次变量:对评价多元能力的培养氛围为优秀的学生有影响的学术支持,对评价多元能力的培养氛围为中等的学生有影响的朋辈交流和平等文化,对评价多元能力的培养氛围为差的学生有影响的师生交流和交往圈子氛围等。另一方面,出现在回归模型中的学生年级、专业学习经历、讨论关注的内容、活动参与、大学生涯发展打算等因素并未出现在决策树分析中。

两种不同的分析工具所展示的不同结果表明了不同的数据观点。回归分析和决策树分析给出的统计差异源于前者是基于最小二乘法的线性技术,后者是基于 χ^2 统计技术。不同的结果反映出了两种技术分

析总体的差异:回归分析将满意度作为一个整体来确定对其影响显著的预测量,而决策树分析的优势之一是对不同分组分别确定不同的解释变量。

本研究基于一所 985 综合性大学的学生数据进行校内比较分析,可以对学术计划的制定进行支持,将学生总体描述为一个不均匀的样本,通过知识经验而不是人口统计将其区分开来。特别的,因为社交系统融入似乎对那些学术参与相对较少的学生较为重要,这个分析暗示了将校园生活与学生融为一体的项目开发的重要性,诸如学习社区建设等可以提升融入社交活动的结构对学习稍差的学生的满意度有更大影响。

仅借助回归分析得不到这些结论,决策树分析通过确定不同的预测变量和学生本身的差异,提供了一种不同的视角。学生本身的差异进一步揭示了高校学

生的不均匀性,这次研究结果表明不同的学术经历作为多样性的一个重要维度的重要性。因此,将回归分析和决策树分析相结合的教育数据挖掘及其结果的学习分析解读,为学生满意度的影响因素确定提供了更宽广的视角。

基于以上分析,为提高学生学习满意度、提升人才培养质量,高校一方面要努力营造多元能力的培养氛围、提供丰富的学习资源、提升教师授课水平,另一方面也要积极提倡并加强师生交流、朋辈交流和交往圈子氛围;为不同类型的学生分别融入学术系统或/和社交系统提供良好支持。上述分析是基于所高校进行的校内分析,通过比较不同大学经历为学生满意度的探究提供了一些见解,但对不同高校的大学经历进行横向比较时,可能会得到不同的结果,因而结论的普适性仍有待进一步测试。

[参考文献]

- [1] Belyukova, S. A., Fox, C. M. Student Satisfaction as A Measure of Student Development: Towards A Universal Metric [J]. Journal of College Student Development, 2002, 43(2):161~172.
- [2] Astin A W. Assessment for Excellence: the Philosophy and Practice of Assessment and Evaluation in Higher Education [M]. Phoenix, AZ: Oryx, 1993.
- [3] Pace, C.R. Quality, Content and Context in the Assessment of Student Learning and Development in College [M]. Washington, DC: Office of Educational Research and Improvement, 1986.
- [4] 傅承哲.本土化学生学习调查调查工具的开发初探[J]. 复旦教育论坛, 2012, 10(3):43~47.
- [5] 祝智庭, 沈德梅. 学习分析学: 智慧教育的科学力量[J]. 电化教育研究, 2013, (5): 5~12, 19.
- [6] Siemens, G. and R.S.J.d. Baker, Learning Analytics and Educational Data Mining: Towards Communication and Collaboration[C]. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, ACM: Vancouver, British Columbia, Canada, 2012:252~254.
- [7] Terenzini, P. T., Pascarella, E. T. Twenty Years of Research on College Students: Lessons for Future Research [J]. Research in Higher Education, 1991, (32): 83~92.
- [8] Tinto, V. Dropout from Higher Education: A Theoretical Synthesis of Recent Research [J]. Review of Educational Research, 1975, (45):89~125.