

■ 人才培养

DOI:10.15998/j.cnki.issn1673-8012.2024.05.008

博士生毕业因素的关联规则挖掘： 群像特征与类型差异



王婷¹, 耿有权², 黄亚继¹

(1. 东南大学 研究生院, 南京 210096; 2. 东南大学 学习科学研究中心, 南京 210096)

摘要:博士生教育以培养具有学术创新能力的高层次人才为目标,肩负着知识生产和促进经济社会发展的重任。在大数据背景下,利用博士生培养过程的教育数据,深度挖掘博士生毕业时间与培养各环节的关联规则,发现博士生延期毕业的主要原因,探析不同时间毕业的博士生群体特征,可为高校加强博士生培养管理、提高博士生毕业率提供科学指导。基于某高校研究生管理大数据系统,获取学生入学信息、导师基础信息、综合素质表现、校园行为情况、科研学习情况5类数据,处理整合形成数据集,并进行特征提取和标签体系构建,采用FP-Growth算法挖掘所有标签与博士生毕业因素的关联规则,最终刻画出按期毕业、延期毕业和超期毕业3类博士生的群体画像,并对比分析3类群像呈现的差异性特征。研究发现:博士生入学方式、课程成绩、开题情况、科研成果数量和完成时间、在读期间是否获得奖学金和荣誉表彰、毕业论文送审时间与博士生毕业时间存在强关联性;博士生入学年龄、入学前院校是不是“双一流”高校、导师年龄、导师任博导年限、导师指导学生人数、校园行为习惯与博士生毕业时间关联性不强;按期毕业和超期毕业博士生分别对应一种典型群体画像,延期毕业博士生包含3种典型画像;3类博士生群体画像在学术志趣、学术基础和学术能力3个方面呈现明显差异。为提升博士生培养质量,提高博士生毕业率,需打破刚性限制与隐性歧视,确保优秀生源选拔的有效性;强化过程监督与分流退出,确保培养过程管理的科学性;重视课程供给与基础培养,保障科研实践衔接的流畅性;呵护博士生学术志趣和身心健康,激发其科研探索 and 创新的积极性。

关键词:博士生教育;博士生群体画像;按期毕业;延期毕业;超期毕业

[中图分类号]G643 [文献标志码]A [文章编号]16738012(2024)05011215

修回日期:20240515

基金项目:江苏省学位与研究生教育教学改革项目“大数据驱动下的研究生学位质量动态监测与风险预警研究”(JGKT23-C004)

作者简介:王婷,女,重庆人,东南大学研究生院助教,硕士,主要从事研究生教育及管理研究;

耿有权,男,安徽全椒人,东南大学学习科学研究中心教授,博士生导师,主要从事世界一流大学建设理论和研究生教育基本理论研究;

黄亚继,男,江苏如皋人,东南大学研究生院教授,博士生导师,主要从事动力工程及工程热物理、学位与研究生教育研究。

引用格式:王婷,耿有权,黄亚继.博士生毕业因素的关联规则挖掘:群像特征与类型差异[J].重庆高教研究,2024,12(5):98112.

Citation format: WANG Ting, GENG Youquan, HUANG Yaji. An exploration of association rules for doctoral graduation factors: group image features and type differences[J]. Chongqing higher education research, 2024, 12(5): 98112.

一、问题提出

当前,中国社会正在向创新型社会转型,博士生作为我国创新体系中重要的生力军,其教育质量和培养数量对我国创新力和竞争力的提高尤为重要^[1]。在从“研究生教育大国”向“研究生教育强国”转型的过程中,从国际发展经验和我国发展实际来看,进一步扩大我国博士生招生数量和培养体量仍然具有一定的必要性^[2]。教育部发展规划司发布的《2023年全国教育事业发展的基本情况》数据显示,2023年招收博士生15.33万人,比上年增长10.29%^[3]。然而,随着博士生招生规模的扩大,博士生延期毕业的现象也日趋严重,能否顺利毕业成为博士生教育中不可忽视的问题。《中华人民共和国高等教育法》规定,博士研究生教育的基本修业年限为3-4年。目前,各高校对博士生学制和最长学习年限的规定大体一致,学制一般为3-4年,最长学习年限为6年。在2017年的相关调查中,博士生延期毕业率为39.7%^[4],2021年“全国博士毕业生调查”数据显示,博士生延期毕业率近50%^[5]。这意味着每年入学的博士生,仅有半数能够在高校规定的学习年限内顺利毕业,更严峻的是,延期毕业率呈现潜在升高趋势。我国博士生教育逐步进入内涵式发展时期,高校更要聚焦博士生培养过程,直面延期毕业博士生教育管理这一现实问题。

博士生科研过程和内容的开放性、挑战性和创新性,决定了博士生毕业时间的不确定性。博士生能否在学制内或在最长学习年限内毕业,需要高校从博士生入学起,持续关注整个培养周期,深入各环节,发现博士生培养过程中存在的问题和不足,分析毕业生延迟毕业的关键影响因素,进而找到相应的解决办法。纵观整个培养过程,究竟哪些因素、哪些环节在影响甚至阻碍博士生按时完成学业?不同时间节点毕业的博士生群体有怎样的特征?这些特征是否具有差异性?这是本研究重点讨论和需要分析解决的问题。教育信息化2.0行动的实施使得高校博士生从入学到毕业全流程的管理数据更加全面,存储更为完整,为分析博士生教育管理规律提供了有力支撑,也为以上问题的解决带来新的思路。

综合上述思考,本研究以某高校整个年级学术型博士生为研究对象,从用户画像研究视角出发,基于该年级完整培养流程大数据,采用数据挖掘方法,分析博士生培养过程与毕业时间的潜在关联规则,探究博士生延期毕业的影响因素,对按期毕业、延期毕业、超期毕业3类博士生群像进行刻画,并对不同时间毕业的博士生画像差异进行分析,以此为优化博士生培养过程提供辅助决策依据。

二、文献综述

(一)关于博士生毕业时间影响因素的研究

博士生毕业意味着完成了博士生教育的培养环节,达到了博士生培养目标。对学术型博士毕业生而言,应当具备知识、能力、品质和谐发展构成的“学者”的基本素质^[6]。因此,不同时间节点毕业的博士生的差异性,究其本质,源于他们在学术基础、学术能力和学术志趣等方面的差异。但学术基础、学术能力和学术志趣难以客观量化,这为研究不同时间毕业博士生特征带来困难。

通过对博士生毕业时间影响因素的相关文献分析可知,博士生毕业时间最主要的影响因素集中在学生个体特征、导师指导情况、高校培养制度3个方面。学生个体特征是影响博士生延期毕业的最主要原因,读博动机、科研能力、学习经历等都会对毕业时间产生影响。积极的、明确的读博动机有助于博士生顺利完成学业,学术动机的功利化和外化的倾向则对博士生的学业完成有负面影响^[7]。在科研能力与学业精力上,积极的学业参与、明确的学业规划、较强的科研能力对毕业时间有显著的正向影响^[8]。在招生入学方式上,硕博连读博士生是否比非硕博连读的博士生如期毕业率更高,并无

统一的结论。但有研究指出,硕博阶段就读于同一学校的博士生延期率比硕博不同校的博士生更高^[9-10]。在其他学生个体特征中,诸如年龄偏大、家庭困扰、财务负担等因素,也会影响博士生按期毕业^[4]。在导师指导情况方面,导师指导博士生人数越少,导师指导经验越丰富,博士生延期毕业概率越低^[9,11]。在高校制度要求方面,影响博士生毕业时间的直接因素是高质量学术论文发表数量,博士生发表高质量学术论文的难度大,耗时长,部分博士生难以在规定学习年限内完成,不得已而延期毕业^[12]。博士生中期考核等过程性监督机制的缺失或缺位,导致部分博士生降低自我约束,学习效率下降,也会导致延期毕业^[13]。延期毕业并非我国博士生教育的特有现象,从20世纪80年代开始,美国、英国、加拿大等国高等教育界开始关注博士生修学时间延长的问题^[14-15]。根据美国国家科学基金会的数据,在美国,博士生毕业所需平均时间为7.5年,相较于1988年的8.7年虽略有下降,但整体上仍处在高位。在欧洲,约有40%至50%的博士生放弃学位^[16]。从国外高校来看,影响博士生修业年限的因素主要包括课程设计、经济资助、院系文化、论文选题、导师指导等方面,个人先赋性因素如年龄、性别、家庭等在博士生毕业时间的预测上是否有效,仍存在争议^[17-19]。

诚然,国内外已有研究成果为本研究提供了坚实的基础,但既有研究仍存在一定局限性。一是研究视角多局限于博士生个体,目前尚未出现对博士生毕业情况的总体探讨;二是研究数据多来自问卷调查,获得的研究数据存在主观性和随机性;三是研究方法多采用回归分析,以各影响因素为自变量,延期毕业为因变量,但延期毕业是多种因素累积和交织影响的结果,缺乏过程性变量和多维度变量关联关系的探讨。借助已有研究,本研究仍将博士生毕业时间影响因素的探究聚焦在个体特征、导师情况、培养过程3个方面,但区别于已有研究,采集某一年级的所有博士生在培养过程中的系统数据,采用数据挖掘方法中的关联规则分析,重点分析博士生培养全流程中涉及的因素与博士生延期毕业的关联规则。

(二)关于教育数据挖掘的研究

教育数据挖掘是一个学科交叉领域,主要是从教育领域产生的具有独特教育性质的数据里发现隐藏规律,从而更好地分析和理解学生的行为模式,寻找提升学生学习效果的方式与方法^[20]。

教育数据挖掘在美国高校中已广泛运用于吸引优质生源、优化课程设置、吸引经费投入等方面^[21]。例如,美国加州大学两所分校通过对4.8万名申请入学学生的来源地区、种族、语言等进行数据挖掘,并对预测入学概率40%以上的学生进行针对性宣传,使得该部分学生入学率从6.4%提升至21.9%^[22]。威斯康星麦迪逊分校为筹集更多校友捐赠,深入分析了1982年以来的捐赠数据,构建校友捐赠预测模型对17万名校友进行挖掘后分类宣传,成功募集到更多教育经费^[23]。在我国,教育数据挖掘主要应用在3个方面。第一,学生家庭经济困难判定,通过挖掘学生的校园消费信息和与之相关的其他校园行为数据,找到内在关联,进而准确判定学生家庭真实经济情况,实现精准资助^[24]。第二,学生学业情况的掌握,主要是通过对学生学习行为、生活习惯、入学信息、在校表现等数据与学业数据进行分析,预测学生的学习成绩、毕业情况,为招生录取、管理培养、就业指导等提供决策依据^[25-26]。第三,学校管理和服务水平的提升,主要通过学生行为数据及后勤系统数据,发现学生行为规律,从而提高管理服务质量。例如,根据学生图书借阅行为来分析学生阅读习惯并进行图书文献精准推荐等^[27]。

自《教育信息化2.0行动计划》实施起,高校在不断优化信息化管理系统的同时,也逐步形成庞大的学生培养和管理数据库。充分利用数据库探寻学生培养过程中留下的“碎片化”数据之间的潜在关系和规律,一方面可实现数据高效运用,另一方面能给高校教育管理带来客观真实的量化认识。为此,本研究为更好地发挥博士生教育管理数据的价值,挖掘“碎片化”博士生培养过程数据之间的

关联性,选择直接通过高校研究生管理系统数据,聚焦博士生培养过程的关键特征因素与博士生毕业时间之间的关联规则挖掘,为博士生毕业“前置式”管理提供支持。

(三)关于学生画像的研究

对于教育数据挖掘结果如何直观清晰地呈现,以便于高校教育管理者理解数据与行为之间的关联这一问题,学生画像带来了新的解决思路^[20]。学生画像是用户画像在教育领域的应用,其方法是通过分析与学生学习相关的数据,对学生进行群体分类并描述,进而为教与学提供数据支撑和服务^[28]。

目前,部分学生画像研究主要基于学习环境,先采集学生的学习背景、成绩数据和平台交互行为数据^[29],再根据学生的学习动机、学习风格、知识水平、学习参与度、学习积极性等特征对学生群体进行划分,以提供个性化指导,提升学习效率^[30]。例如,有学者以图书情报硕士生为研究样本,利用学生论文发表数据,处理形成用户维度和行为维度的用户画像标签数据集,分析群体的潜在科研需求和发文偏好,提供差异化的科研发展策略^[31]。有研究通过收集无线网轨迹数据、智能卡和借阅数据等行为数据,构建学习成绩分类模型并进行用户画像,找到了成绩和各参数之间的规律,提高了学生学习效率^[32]。也有研究根据学生的基本属性、学习过程和学习结果数据构建画像模型,并基于画像结果提出学习路径规划框架建议^[33]。

结合大数据挖掘用户画像,最早且广泛地应用于互联网公司的是基于用户浏览习惯的商品推荐。目前,尚未有研究以博士生群体为目标对象,将用户画像与数据挖掘技术结合,对博士生毕业时间和博士生培养全过程进行关联分析和画像描绘。综上,本研究将从用户画像视角出发,根据博士生毕业时间,将其分为按期毕业、延期毕业和超期毕业3类,使用数据挖掘方法探讨博士生毕业时间与培养过程之间的潜在关系。

三、研究设计

(一)研究思路

研究从用户画像视角出发,聚焦按期毕业、延期毕业和超期毕业3类不同时间毕业博士生的群体画像刻画及成像差异性分析,本质上是对博士生培养过程数据的整合、挖掘和表达。构建画像的基础是对研究对象特征的梳理和筛选,核心是从大量数据中提取研究对象的典型标签。对本研究而言,就是通过数据挖掘的方法分析博士生培养过程中哪些特征因素可能与其毕业时间存在强关联。

第一步,研究对象数据化,即集合形成以学生入学信息、导师基础信息、综合素质表现、校园行为数据和科研学习情况为一体的目标数据,通过从海量的博士生培养过程数据中提取有分析价值的数
据,构建标签体系,形成研究数据集。第二步,研究对象特征化,即通过构建基于FP-Growth算法的博士生毕业关联规则挖掘模型,基于关联规则挖掘数据集中所有因素标签与博士生毕业时间的关联关系,实现强关联规则输出,并描绘3种不同时间毕业的博士生群体画像。第三步,通过对比分析不同群体画像之间的显著差异,提出具有可操作性的建议,旨在为博士生培养提供有力的决策支持。具体研究路线如图1。

(二)画像构建方法

对不同时间毕业博士生进行画像构建,是通过对不同
类型博士生群体的基本信息和培养过程大数据进行群体共同特征提取,进而抽象出标签化的形象模型,即对典型群体的标签化和模型化。目前运用于画像构建的方法主要有3种,包括定性用户画像方法、基于定性研究的定量用户画像方法和经定量检验的定性用户画像方法^[34]。其中,定性画像的核心工作是为不同群体打标签,刻画并呈现其群体共性特征。因本研究采用的FP-Growth算法通过频繁项集直接输出相关性最高的静态、动态标

签和各标签的强关联关系,故本研究采用定性用户画像方法对不同时间毕业的博士生进行画像构建。

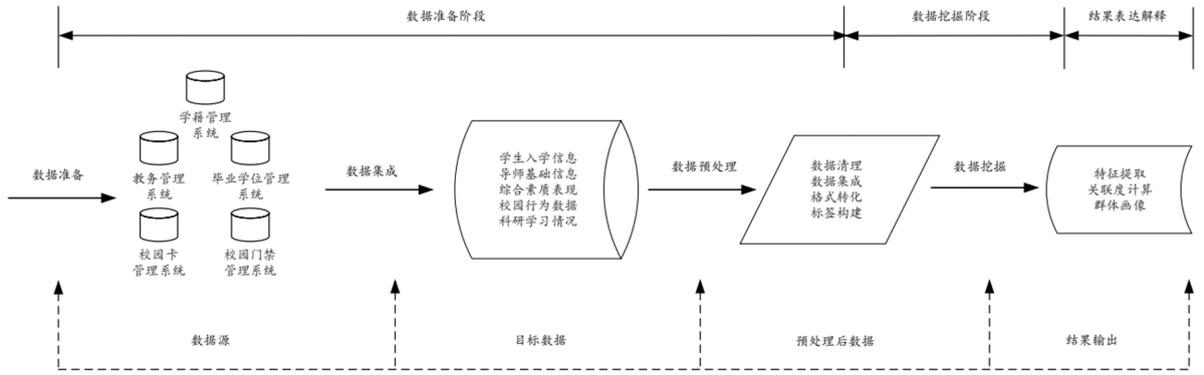


图1 研究路线

(三)数据准备

1. 样本选取

本研究选取国内某“双一流”建设高校 2017 年入学的学术型博士生为研究对象,为确保样本间的可比性,将培养要求存在差异的直博生、专业学位博士生排除,最终纳入分析的样本量共计 652 人。根据该高校研究生学籍管理规定,博士生学制为 4 年,最长学习年限为 6 年,超过学制但未超过最长学习年限即为延期,超过最长学习年限即为超期。因此,本文以上述两个时间节点将研究对象分为 3 类群体,分别是:按期毕业博士生,占比 24%;延期毕业博士生,占比 56%;超期毕业博士生,占比 20%。

2. 数据预处理

基于已有研究,本研究从学生个体特征、导师指导情况、学习培养过程 3 个维度进行讨论,从博士生招生录取、培养管理、学位申请的全过程选取学生入学信息、导师基础信息、学生在校综合素质表现、校园行为情况、科研学习情况 5 个方面 15 个因素进行数据收集(如图 2)。该高校博士生培养过程所有数据均以信息系统形式存储,本研究的数据从该高校的研究生学籍管理系统、教务管理系统、毕业与学位管理系统、校园一卡通使用管理系统、校园门禁管理系统等直接获取。因涉及学生个人隐私,所有数据都经过脱敏和匿名化处理。为提升数据质量,确保模型分析的效果,建模前本研究对所有数据均进行数据清理、数据集成和格式转化等数据预处理,所有数值型数据均进行归一化处理,经数据预处理后共计 4 万余条。

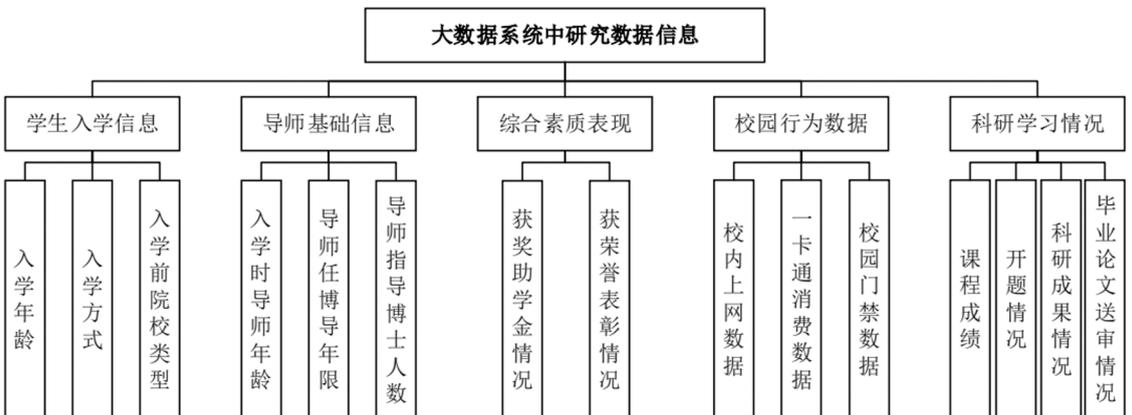


图2 研究因素框架

3. 标签体系构建

标签体系是对不同标签进行科学归类,并对标签属性进行定义从而形成体系,此过程是为了便于规范标签的管理和使用。本研究中 15 个因素涉及的 44 个标签,根据属性可分为静态表征标签和动态表征标签。静态表征标签包括学生入学信息、导师基础信息 2 个影响因素覆盖的 16 个标签,动态表征标签包括毕业时间情况和综合素质表现、校园行为数据、科研学习情况 3 个影响因素覆盖的 28 个标签。最终建立的标签体系及对应符号如表 1 和表 2 所示。

表 1 静态标签符号及其含义

静态标签	标签含义	静态标签	标签含义
B1	入学年龄—25 岁前	E1-1	入学时导师年龄低于 40 岁
B2	入学年龄—25~30 岁	E1-2	入学时导师年龄 40~50 岁
B3	入学年龄—30 岁后	E1-3	入学时导师年龄 50 岁以上
C1	入学方式—“申请—考核”	E2-1	入学时导师任博导不足 5 年
C2	入学方式—硕博连读	E2-2	入学时导师任博导 5~10 年
D1	硕士毕业院校—“双一流”高校	E2-3	入学时导师任博导 10 年以上
D2	硕士毕业院校—非“双一流”高校	E3-1	毕业时导师指导博士人数小于 5 人
E3-3	毕业时导师指导博士人数大于 10 人	E3-2	毕业时导师指导博士人数 5~10 人

表 2 动态标签符号及其含义

动态标签	标签含义	动态标签	标签含义
A1	按期毕业	H1-1	课程规格化成绩—低于 80 分
A2	延期毕业	H1-2	课程规格化成绩—高于 80 分
A3	超期毕业	H2-1	按期开题
F1	获得过奖助学金或荣誉表彰	H2-2	延期开题
F2	未获得过奖助学金或荣誉表彰	H2-3	未开题
G1-1	每日平均上网时长小于 5 小时	H3-1	科研成果数量超过毕业要求
G1-2	每日平均上网时长 5~8 小时	H3-2	科研成果数量达到毕业要求
G1-3	每日平均上网时常超过 8 小时	H3-3	科研成果数量未达到毕业要求
G2-1	每日食堂消费小于 3 次	H4-1	科研成果发表达标时间—学制内
G2-2	每日食堂消费 3~5 次	H4-2	科研成果发表达标时间—学习年限内
G2-3	每日食堂消费超过 5 次	H4-3	科研成果发表达标时间—超学习年限
G3-1	每日宿舍停留时间小于 8 小时	H5-1	毕业论文送审时间—学制内
G3-2	每日宿舍停留时间 8~12 小时	H5-2	毕业论文送审时间—学习年限内
G3-3	每日宿舍停留时间超过 12 小时	H5-3	毕业论文送审时间—超学习年限

(四) 数据挖掘

1. 挖掘算法选择

本研究的数据挖掘重点在于关联规则分析,关联规则作为数据挖掘技术的重要分支,可以捕捉数据之间的潜在关联。为提高构建效率,加强生成频繁项集相关性,本研究采用改进的 FP-Growth 算法挖掘培养过程中各因素与博士生毕业时间的关联规则^[35]。改进算法增加了剪枝过程,即增加了对每个样本分支中是否含有博士生毕业情况频繁项的判断条件,若有就保留该分支,若没有就剪掉该分

支。在遍历 FP-Tree 生成频繁项集时,也增加对频繁项集中是否含有博士毕业情况频繁项的判断,如果有就保留频繁项集,如果没有则删掉该频繁项集,具体数据挖掘流程如图 3 所示。

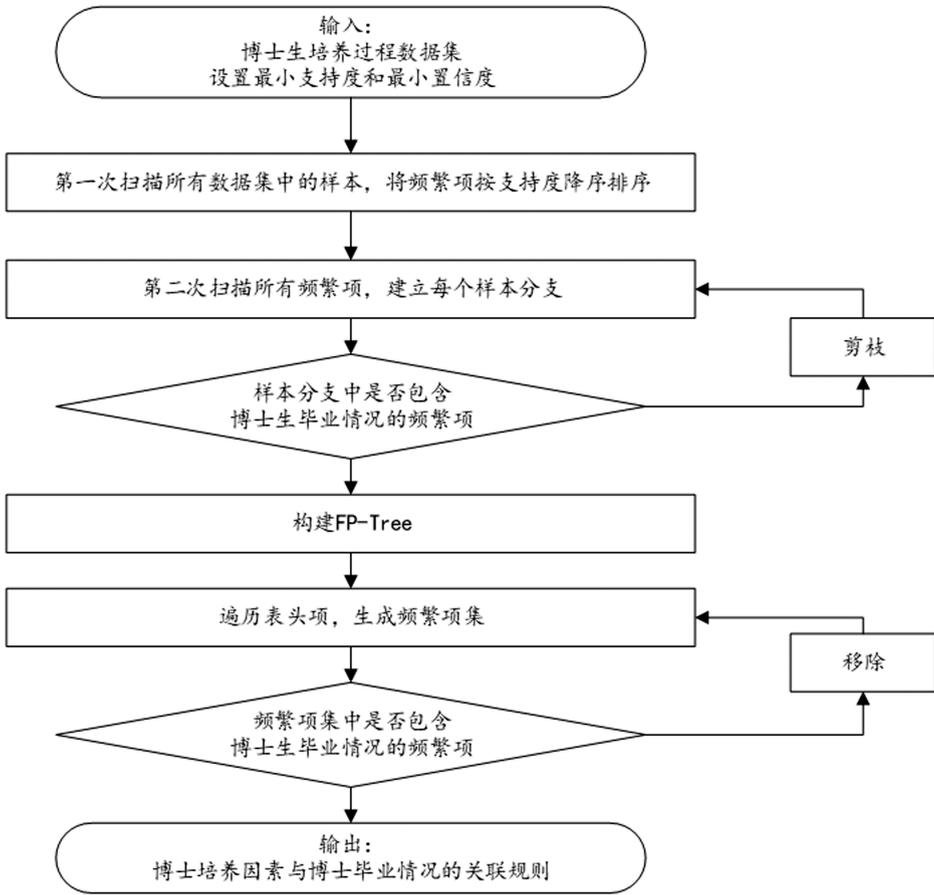


图 3 基于 FP-Growth 算法的博士生毕业情况数据挖掘流程

2. 关联规则挖掘

关联规则定义为 $X \rightarrow Y$ 形式的蕴含表达式,其中 X 和 Y 是不相交的项集,即 $X \cap Y = \emptyset$ 。关联规则表示不同项目集合之间的相互依存性和关联性。关联规则的强度通过支持度、置信度和提升度来衡量。

支持度 (support), 表示某个项集在数据库中出现的概率,即 X 和 Y 同时出现的概率:

$$s(X \rightarrow Y) = (\sigma(X \cup Y)) / N$$

置信度 (confidence), 表示 X 和 Y 同时出现的概率占 X 出现概率的比值:

$$c(X \rightarrow Y) = (\sigma(X \cup Y)) / (\sigma(X))$$

置信度用于衡量规则的强弱程度,置信度越高,说明 X 出现时, Y 出现的可能性越大。

提升度 (lift), 表示在 X 出现的条件下出现 Y 的概率和没有条件 X 出现 Y 的概率:

$$l(X \rightarrow Y) = (\sigma(X \cup Y)) / (\sigma(X) \times \sigma(Y))$$

提升度直接反映关联规则中 X 与 Y 的相关性,提升度 > 1 且越高表明正相关性越高,提升度 < 1 且越低表明负相关性越高,提升度 $= 1$ 表明没有相关性。

研究时设置最小支持度和最小置信度阈值,最小支持度表示项集在统计意义上的最低重要性,最小置信度表示关联规则的最低可靠性。当某个关联规则提升度大于 1,且满足最小支持度阈值和最小置信度阈值,则认为该关联规则有效且为强关联规则。

四、结果与讨论

(一)关联规则输出结果

在进行基于 FP-Growth 算法的数据挖掘之前,本研究首先将不同毕业时间的博士生进行分类,再分别进行数据集的差异和相似分析。本研究最小支持度阈值和最小置信度阈值分别设置为 0.2 和 0.3。根据支持度、置信度和提升度的计算,输出关联规则的支持度和置信度都大于阈值,且提升度大于 1 的关联规则,即 3 类博士生群体对应的毕业影响因素强关联规则如表 3 所示。

表 3 博士生毕业时间强关联规则输出结果

毕业时间	关联规则	支持度	置信度	提升度
按期毕业	{H3-1, C2, F1, H5-1, H4-1} ⇒ A1	0.315	0.526	1.863
	{H3-1, C2, H4-1, F1, H5-2} ⇒ A2	0.289	0.402	1.814
延期毕业	{H1-2, F1, H4-2, H5-2} ⇒ A2	0.284	0.391	1.423
	{H4-1, F2, C1, H3-2, H2-2} ⇒ A2	0.278	0.415	1.516
超期毕业	{F2, C1, H5-3, H4-3, H3-3, H2-2} ⇒ A3	0.332	0.564	2.056

表 3 结果表明:

第一,博士生入学方式(C)、博士在读期间是否获得奖学金和荣誉表彰(F)、课程成绩(H1)、开题情况(H2)、科研成果发表的时间和数量(H3、H4)、毕业论文送审情况(H5)这 7 个因素与博士生毕业时间存在强关联性。

第二,博士生入学年龄(B)、入学前院校是不是“双一流”高校(D)、导师年龄(E1)、导师任博导年限(E2)、导师指导学生人数(E3)、博士生校园行为习惯(G1、G2、G3)共计 8 个因素与博士生毕业时间无强关联性。

第三,根据关联规则理论,每条强关联规则代表着一种典型的特征组合,即代表一种典型人群。按期毕业博士生和超期毕业博士生分别有一种典型群体,而延期毕业博士生则包含了 3 种不同典型群体。

(二)不同博士生群像特征分析

1. 按期毕业博士生群像特征分析

按期毕业强关联规则代表博士生通过硕博连读方式入学,在学制内完成科研成果发表并完成毕业论文送审,科研成果数量超过毕业要求,在博士阶段获得过奖学金或荣誉表彰。首先,此类博士生严格按照学校培养要求的时间节点完成每项培养计划,在学制内能按时完成科研成果发表、毕业论文送审等培养要求,具有明确的毕业目标倾向和完整的学业生涯规划。其次,他们在学制内能取得超出毕业基本要求的科研成果,获得奖学金或荣誉表彰,说明此类博士生学术能力较为突出,属于综合表现优异的博士生群体。此外,学制内毕业的博士生大多通过硕博连读方式入学,博士入学前已接受了较为系统的学术训练,具备一定的学术研究基础和实力,在学术能力、科研产出等方面均展现出硕博一体化培养的优势^[10]。综合看,这类博士生特征可概括为:硕博连读入学,学业规划明确,科研学术能力强,学术成果产出速度快、数量多,综合表现突出。

2. 延期毕业博士生群像特征分析

相较于按期毕业的博士生和超期毕业的博士生强关联规则的唯一性,延期毕业博士生强关联规则有 3 条,即延期毕业博士生包含 3 种不同群体。

第一种延期毕业强关联规则代表博士生通过硕博连读方式入学,在学制内完成毕业要求的科研成果,且科研成果数量超过毕业基本要求,获得过奖学金和荣誉,但毕业论文送审时间超期。此类延期毕业博士生与按期毕业博士生入学方式相同,均为硕博连读入学。从硕士阶段转入博士阶段后,在学制规定时间内科研成果数量丰硕,证明了其较高的科研素质、较强的学术能力和有序的学业规划。博士在读期间获得奖学金和荣誉表彰,也是其学术能力优秀的表现。但该群体往往因毕业论文送审时间推迟导致毕业延期。综合博士生培养实际过程和已有研究结果,分析此类博士生毕业论文延迟送审的原因有3种可能。第一种可能是到规定学制时间,此类博士生虽然已满足毕业条件,但他们仍需要对已有学术成果进行更深入研究,花费了更长的时间在取得创造性、突破性的学术研究成果上,影响了毕业论文的送审时间。第二种可能是在高年级时他们选择公派出国等联合培养项目,在开拓学术视野、增强国际眼界等方面花费了精力和时间,进而主动选择推迟毕业时间^[36]。第三种可能是因突出的科研能力,这类博士生在课题组科研攻关中起到骨干作用,是导师项目的主力军,承担科研任务的周期存在不确定性,无形中延长了博士生毕业的时间^[37]。不管源于哪种可能性,培养时间的延长使得此类博士生增加了理论知识积累,提升了科研能力水平,出现延期毕业是进一步深入学术研究的合理表现。综合看,这类博士生群体特征可概括为:硕博连读入学,学术兴趣强烈,科研能力强,学术成果产出速度快、数量多,综合表现突出。

第二种延期毕业强关联规则代表博士生课程规格化成绩高于规格化平均水平,博士阶段荣获奖学金和荣誉表彰,在学制内完成科研成果要求,毕业论文送审时间超过学制但在最长学习年限内。此类博士生从小的应试教育让他们非常顺利地度过课程学习阶段,甚至能取得优异的课程成绩。他们从课程学习到科研探索均按照博士生培养计划各项培养要求进行,取得的科研成果虽然数量不突出但学术论文的发表都在规定时间内,对自己博士阶段学业规划有整体性把握。他们综合表现良好,获得过奖学金和荣誉表彰。但是,博士毕业学位论文作为博士生学习和研究的最终成果,不是已有科研成果的拼凑,而是学科基础、科研能力、创新能力、掌握和运用知识能力以及书面表达能力等的综合体现^[38]。此类博士生在博士毕业论文研究后期,暴露出知识储备和科研能力不足,论文写作和送审出现停滞或滞后。综合看,这类博士生特征可概括为:课程成绩优异,学业规划明确,科研学术水平能达到毕业要求但不突出,毕业论文送审延迟。

第三种延期毕业强关联规则代表博士生以“申请—考核”方式入学,开题时间出现延迟,能在规定学习年限内完成规定科研成果,但成果数量刚达到毕业要求,在博士就读期间未获得过奖学金或荣誉表彰。此类博士生通过“申请—考核”入学,开题出现了延迟,培养初期就表现出学术基础薄弱、科研能力不足等情况。但经过2-3年的学术训练后,他们在学术探索阶段科研水平得到提升,顺利打破了论文选题与科研能力失衡的现状,提升了自身科研能力阈限,在规定学习年限内发表了规定数量要求的科研成果,最终达到博士生毕业要求,顺利完成学业。综合看,这类博士生特征可概括为:“申请—考核”入学,基础薄弱,开题延迟,科研学术水平仅能达到基本毕业要求。

3. 超期毕业博士生群像特征分析

超期毕业强关联规则代表博士生通过“申请—考核”方式入学,开题延期,科研成果发表数量不满足要求,科研成果发表时间超过最长学习年限,博士在读期间未获得过奖学金或荣誉表彰,毕业论文送审时间超过最长学习年限。超最长学习年限还未毕业的博士生群体,入学后因为学术基础薄弱,普遍从开题环节就出现滞后,过渡适应阶段就出现了适应性障碍。随着培养时间的推进,科研成果发表、毕业论文完成时间也随之滞后,除课程学习外的其他培养环节,均无法按要求按时完成,最终导致无法在规定时间内毕业。此类博士生群体入学前已经存在学术基础差、学术能力弱的情况,步入博士

培养阶段后,所具备的知识储备、科研能力与实际学术过程存在偏差,且容易在课题研究过程中出现畏难情绪和行为,丧失对学术研究的兴趣。在博士生评奖评优过程中,此类学生也较难获得奖学金或荣誉表彰。综合看,这类博士生特征可概括为:以“申请—考核”入学,从开题环节起每个培养环节都滞后,学术基础差,科研能力弱。

综上,3类博士生画像的主要特征概括如表4所示。

表4 3类博士生画像主要特征

类型	主要特征
按期毕业	硕博连读入学,学业规划明确,科研学术能力强,学术成果产出速度快、数量多,综合表现突出
延期毕业Ⅰ	硕博连读入学,学术兴趣强烈,科研能力强,学术成果产出速度快、数量多,综合表现突出
延期毕业Ⅱ	课程成绩优异,学业规划明确,科研学术水平能达到毕业要求但不突出,毕业论文送审延迟
延期毕业Ⅲ	“申请—考核”入学,基础薄弱,开题延迟,科研学术水平仅能达到基本毕业要求
超期毕业	“申请—考核”入学,从开题起每个培养环节都滞后,学术基础差,科研能力弱

(三)不同博士生成像差异及讨论

学术型博士毕业生应具有浓厚的学术志趣、坚实的学术基础和卓越的学术能力3个基本素质^[6,39]。进一步分析群像特征发现,与博士生毕业时间强关联的7种因素通过作用于博士生的学术志趣、学术基础和学术能力而影响博士生的毕业时间。其中,博士生入学方式的不同、课程成绩的好坏反映了博士生学术基础和学术能力的差异;博士生在读期间是否获奖学金和荣誉表彰、是否按期完成开题、科研成果发表时间和数量,以及毕业论文送审时间反映了博士生的学术志趣、学术基础和学术能力的差异。

1. 饱满的学术志趣:克服困难的动力

学术志趣立足兴趣,着眼志向,引领博士生探索创新,引导博士生增强学业规划的自觉意识。博士生在入学时都充满“雄心壮志”,但当学术热情与科研实际之间产生偏差时,不同博士生的学术志趣将出现差异性转变。从画像特征看,按期毕业博士生和第一类延期毕业博士生,在博士就读期间学业规划明确,学术兴趣强烈且呈现加强的趋势,浓厚的学术志趣支撑他们全身心投入科研中,严格按照按时完成每个培养环节要求。在延期毕业的博士生中,大部分入学时有较明确的目标志向和浓厚的科研兴趣,且能在读博前期保持住,但学术志趣在学术成果发表、毕业论文撰写等不同时间节点出现减退,导致延期。超期毕业的博士生普遍从开题阶段学术志趣就逐渐减退,且消极被动的学习状态和意志力的减弱很难支撑他们顺利完成学业。

2. 坚实的学术基础:支撑探索的基石

学术基础是博士生开展博士阶段学术研究的基本依托,是形成和发展提出问题、分析问题和解决问题能力的必要保障^[40]。博士生入学前如果自身专业基础薄弱,寄希望于入学后通过课程或科研训练快速养成是比较困难的。从按期毕业和延期毕业博士生群像看,硕博连读学生在学术基础方面呈现一定优越性,他们因学术经历的连贯性,提前进入课题组,博士入学前积累了一定的学术基础,博士培养阶段的科研道路相对顺利。从超期毕业博士生群体入学后的表现来看,部分“申请—考核”入学的博士生暴露出了学术基础薄弱的问题。第二类延期毕业博士生作为应试型学习者,虽具备一定的学科基础,但博士生培养要求在专业领域内掌握坚实的基础理论和系统深入的专门知识,而这仅靠应试的知识和能力是远远不够的。

3. 卓越的学术能力:创新研究的核心

学术能力是衡量博士生培养质量的基本指标,指博士生在学习、研究、技术开发和社会实践的过程中形成的,并通过完成论文、承担课题等方式将之付诸实践的能力^[41]。从3类群体画像看,按期毕业博士生及第一类延期毕业博士生,在课题研究推进、学术成果产出、毕业论文撰写方面都有更为突出的表现,按期毕业是水到渠成之事。后两类延期毕业博士生,科研综合能力在入学时较弱,在后期的课题研究、论文发表等学术锻炼中,学术能力得到提升,最终顺利毕业。超期毕业博士生从入学时就暴露出学术能力欠缺的问题,从博士论文开题起各培养环节难以在规定时间内完成,而学术能力又难以在短时间内有效提升。

基于以上分析,将5类画像在学术志趣、学术基础和学术能力3个方面的差异性集中呈现,如图5所示。

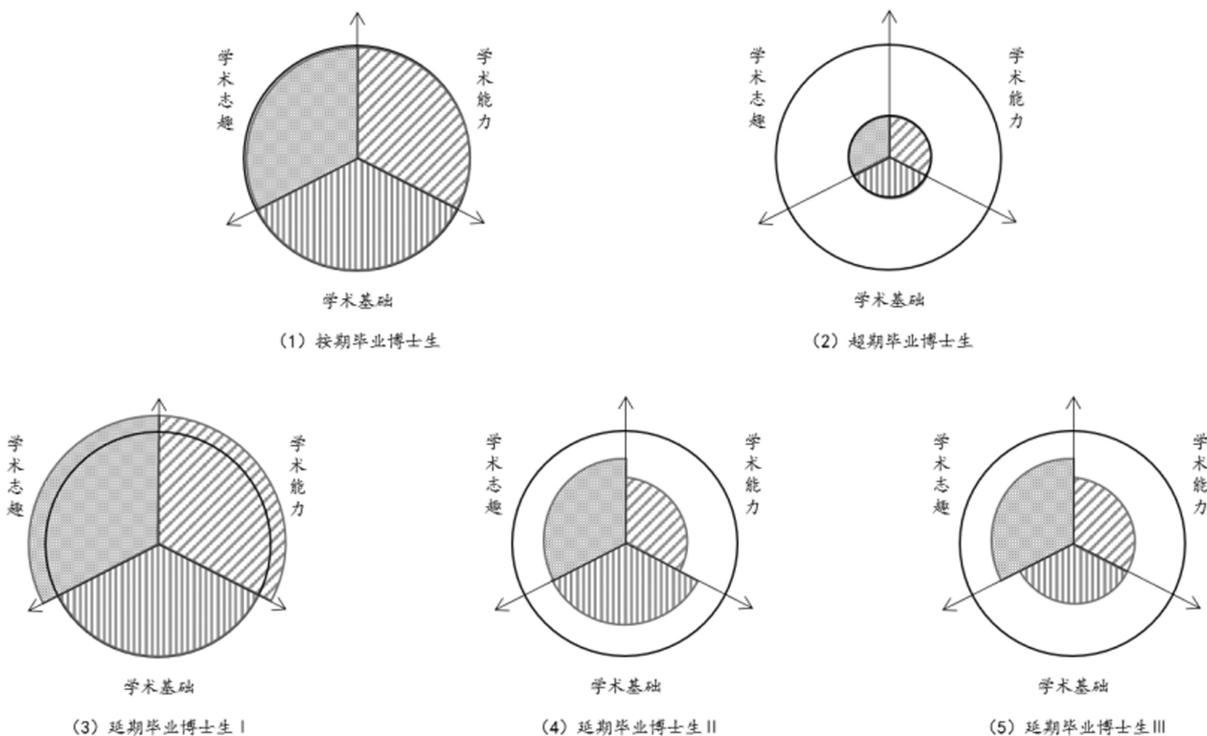


图5 5类群像的差异性显示

五、结论与建议

(一) 研究结论

本研究对国内某“双一流”建设高校博士生的学生入学信息、导师基础信息、综合素质表现、校园行为情况、科研学习情况五大类数据进行数据挖掘和群像分析,得出以下主要结论:

1. 影响因素:入学方式和培养过程表现影响显著,人口学特征因素、前置院校类型和导师客观信息因素影响不显著

在本研究的15个影响因素中,博士生入学方式、课程成绩、开题情况、科研成果数量和完成时间、在读期间是否获得奖学金和荣誉表彰及毕业论文送审时间与博士生毕业时间存在强关联性。这一结论进一步说明,博士生毕业时间与博士生每个培养环节能否顺利通关的关联性最大,该结论在已有文献中可得到印证^[42]。博士生入学年龄、入学前院校是不是“双一流”高校、导师年龄、导师任博导年限、导师指导学生人数、校园行为习惯等8个因素与博士生毕业时间关联性不强,该研究结论与以往

研究结论有差异^[4,9-10],这种差异可能源于研究方法和调查样本的不同,可在今后研究中进一步证实。

2. 画像特征:按期毕业、延期毕业和超期毕业博士生分别具有明显不同的画像特征

通过对不同博士生群体画像特征的分析发现,不同博士生群体确实存在不同的典型群体画像特征,可以为预测和识别按期毕业、延期毕业、超期毕业博士生提供参照。对比发现,按期毕业博士生和延期毕业中第一类博士生群体,培养成长过程更加契合博士生培养规律和要求,这两类学生画像特征可为其他延期毕业情况的研究提供参考。后两类延期毕业博士生和超期毕业博士生,因学业困难无法按期毕业,是高校应该重点关注和帮扶的学生群体。

3. 成像差异:3类博士生在学术志趣、学术基础和学术能力3个方面存在差异

综合3类博士生画像特征和不同画像呈现的差异性发现,按期毕业博士生、延期毕业博士生和超期毕业博士生在学术志趣、学术基础和学术能力3个方面存在较大差异。按期毕业博士生和第一类延期毕业博士生在学术志趣、学术基础和学术能力方面都显示出明显优势。相反,超期毕业博士生在这3个方面均存在明显不足。延期毕业的另两类博士生不同画像之间存在差异,但未在3个方面同时存在短板。

(二) 研究建议

1. 打破刚性限制与隐性歧视,确保优秀生源选拔的有效性

对于博士生教育而言,生源质量的重要性不亚于培养质量。随着招生规模的不断扩大,如何在数量增加的同时保证质量的稳定与提升,无疑成为博士招生的难题。本研究发现,博士生的院校出身与毕业时间并不具备明显关联性,“名校出身”不等于“优质生源”,在确定博士申请资格条件时,要摒弃唯出身论,客观评价每一位考生。要重点考查学生学术志趣,学术志趣才是攻读博士学位最持久的内驱力,缺乏学术志趣的学生难以克服读博过程中巨大的科研困难和压力,学术志趣浓厚的学生有更大的学术投入和抱负。在本研究中,硕博连读博士生按期毕业率略高于“申请—考核”的博士生,这并非指哪种招生方式更好,而是提醒高校在博士招生过程中,更应注重对申请者学术基础和学术能力的考察,如果申请者没有扎实的学科和专业功底,就难以胜任繁重的学业要求和科研任务。贯通式培养相比于阶段式培养具备一定优势,例如培养周期短,导师与学生经过硕士阶段的磨合后,能减少培养过程中导生矛盾,但博士生“申请—考核”制招生作为我国博士招生制度改革的重要探索,高校仍需继续探索如何设计科学合理的考核内容和考核形式,确保人才选拔的有效性。

2. 强化过程监督与分流退出,确保培养过程管理的科学性

在本研究样本中两成博士生从入学起就出现了不同程度的培养环节滞后,加强博士生培养的过程性管理迫在眉睫。首先,要将管理重心前移,严把重点关卡。在本研究中,按期开题的博士生,按期毕业的概率越大;反之,毕业论文送审延期,延期毕业概率更大。博士生按期毕业与成功“通关”在一定程度上存在关联,论文开题关通过意味着获得“施工许可证”,论文送审关通过是争取“验收合格证”,验收不通过将直接导致延期,开工延迟还有补救可能,因此,导师应在开题时对选题、思路及进度安排重点把关。第二,要利用信息化手段,建立过程监督管理机制。这一机制需要面向博士生、导师和管理者,还要深入各个环节,确保在各项工作开始前进行预告,并且在出现异常时及时发出预警。同时,要畅通分流选择渠道,健全分流退出制度。国家层面早已明确了“分流退出制度”对于博士生教育制度化建设与体系化保障的重要作用,但效果却不尽如人意。延期毕业率的攀升,仅依靠时间的宽限而未及时止损,会带来学校教育资源和导师指导精力的浪费,高校必须完善分流退出制度,并强化分流退出制度的操作与执行。

3. 重视课程供给与基础培养,保障科研实践衔接的流畅性

高质量课程学习可以强化研究生的科学方法训练和学术素养培养。但从现实情况来看,博士生课程是博士生培养过程中较为薄弱的环节,博士生们也并不认同课程教学的重要性。对比本研究的3类博士生,超期毕业和部分延期毕业的博士生入学时学术基础薄弱,导致科研训练过程受阻,进而影响毕业。通过课程教学强化学术基础,是帮助博士生顺利衔接科研实践的重要方法。高校应重视博士生课程教学改革,在课程内容方面,结合科研实际,跟进学术前沿,加强科研技能、研究方法、实验操作、学术写作类实用性课程。在课程内容和授课方式的选择上,满足学生个性化需求,提高学生自主性,通过对课程教学进行从内容到形式的多元化、个性化改革,强化博士生知识积累,为博士生独立从事科研工作奠定坚实基础。在培养理念上,纠正重科研训练、轻课程教学的错误导向,让博士生和导师共同重视课程教学,强化博士生在学术基础方面的积累。

4. 呵护学术志趣和身心健康,激发科研创新探索的积极性

社会高期望、科研高强度、毕业高压等交织形成巨大的压力网笼罩着博士生群体,必然导致博士生群体的学术志趣减弱,焦虑情绪增加^[43]。Nature 杂志 2019 年对 690 名中国博士生进行调查,结果发现,40% 受访者曾因读博期间的抑郁和焦虑寻求过帮助,该数据比其他国家博士生高出 4%^[44]。在本研究及相关研究中,学术志趣浓厚的博士生能以更积极的方式参与学术活动,进而提升学术表现和缩短毕业时间。高校应探索多元发展性教育质量评价,淡化论文发表数量的评价导向,避免形成“内卷”生态,保护博士生的学术兴趣和学术热情。导师作为博士生培养的第一责任人,要主动关心博士生学习规划、学术进展、职业发展、人际关系、身心健康,主动营造有利于导生相处的课题组文化环境,形成开放包容、团结合作的课题组学术交流氛围,激发和保护博士生学术兴趣。导师在指导博士生时,应坚持因材施教的原则,根据学生的学术志趣确定研究方向,根据学生的学术能力匹配研究课题,并在课题确定后,强化对学生的指导。对于存在指导工作不投入、大量博士生超期未能毕业的导师,应视情节予以相应惩罚。

参考文献:

- [1] 蔡文伯,向燕. 博士规模、学术生产与技术创新关系的门槛效应[J]. 重庆高教研究,2023,11(2):8898.
- [2] 洪大用. 贯彻落实党的二十大精神 加快建设研究生教育强国[J]. 学位与研究生教育,2023(9):47.
- [3] 2023 年全国教育事业发展基本情况[EB/OL]. (20240301)[20240409]. http://www.moe.gov.cn/fbh/live/2024/55831/sfcl/202403/t20240301_1117517.html.
- [4] 鲍威,张心悦,吴嘉琦. 博士生延期毕业影响因素的实证研究[J]. 江苏高教,2020(7):3438.
- [5] 许丹东,沈文钦,翟月,等. 中国博士生的培养现状与问题:基于 2021 年全国博士毕业生离校反馈调查的分析[J]. 学位与研究生教育,2022(5):7380.
- [6] 包水梅. 学术型博士生培养目标定位及其素质结构研究[J]. 教育科学,2015,31(2):7478.
- [7] 蔡芬,曹延飞,顾晔,等. 教育博士生延期毕业影响因素的质性研究[J]. 学位与研究生教育,2020(3):4652.
- [8] 刘玮. 延期毕业博士生的主要特征研究:基于某重点高校数据分析[J]. 中国青年研究,2016(1):4448,74.
- [9] 李海生. 博士研究生延期完成学业的影响因素分析[J]. 复旦教育论坛,2019,17(3):5259.
- [10] 郭海燕,刘春荣,张志斌. 生源差异如何影响博士学位论文质量:基于全数据的研究[J]. 研究生教育研究,2019(2):4349.
- [11] 绳丽惠. 博士生延期毕业现象:影响因素与治理策略[J]. 学位与研究生教育,2019(6):6064.
- [12] 李兰,谢永生. 博士生申请学位时要求发表学术论文的问卷调查与分析[J]. 学位与研究生教育,2015(6):60-64.
- [13] 杨青. 博士生为什么延期了:人文社科博士生延期毕业原因及作用机制分析[J]. 中国高教研究,2021(7):90-97.
- [14] BOURK S, HOLBROOK A, LOVAT T, et al. Attrition, completion and completion times of PhD candidates[J].

AARE annual conference, 2004,28: +14.

- [15] 杨频萍,汪霞. 博士生为什么延期:国外博士生毕业时间影响因素研究[J]. 江苏高教,2021(4):3239.
- [16] GEVEN K, SKOPEK J, TRIVENTI M. How to increase PhD completion rates? an impact evaluation of two reforms in a selective graduate school,19762012[J]. Research in higher education, 2018,59(5): 529552.
- [17] 赵世奎,沈文钦,张帅. 博士修业年限及其影响因素分析:基于中美比较的视角[J]. 教育学术月刊,2010(4): 34-37.
- [18] LIN E S, CHIU S Y. Can a doctoral degree be completed faster overseas than domestically? evidence from Taiwanese doctorates[J]. Higher education,2014,68(2):263283.
- [19] SHARIFF N M, RAMLI K I, AHMAD R, et al. Factors contributing to the timely completion of PhD at the Malaysian public higher educational institutions[J]. International journal of humanities social and education,2015(2):256263.
- [20] 杨孟辉,任超,谢宝玲,等. 基于校园行为特征的学生就业画像[J]. 图书馆论坛,2024,44(1):6270.
- [21] 丁国勇,程晋宽. 美国高校教育数据挖掘的应用领域与推动因素[J]. 比较教育研究,2017,39(5):7378.
- [22] 常桐善. 数据挖掘技术在美国院校研究中的应用[J]. 复旦教育论坛,2009,7(2):7279.
- [23] DAVID J W, JUSTIN M R. Using classification trees to predict alumni giving for higher education[J]. Education economics,2009,17(1):95122.
- [24] 费小丹,董新科,张晖. 基于校园一卡通消费数据的高校贫困生分析[J]. 电脑知识与技术,2014,10(20):4934-4936.
- [25] 钱增瑾,孙东平. 数据挖掘在研究生教育管理信息系统中的应用[J]. 学位与研究生教育,2013(5):4649.
- [26] 卢铮松. 研究生奖学金的决策树分类数据挖掘研究[J]. 计算机工程与应用,2012,48(26):139143.
- [27] 柳益君,何胜,冯新翎,等. 大数据挖掘在高校图书馆个性化服务中应用研究[J]. 图书馆工作与研究,2017(5): 23-29.
- [28] 张兴旭,邢磊. 指向高质量学习的大学生学习者画像构建:以上海交通大学为案例的考察[J]. 教育发展研究, 2023,43(17):7684.
- [29] 张雪,檀悦颖,罗恒. 在线学习非母语学习者群体研究:类别画像与行为特征分析[J]. 现代远距离教育,2019 (1):1826.
- [30] 唐焯伟,茹丽娜,范佳荣,等. 基于学习者画像建模的个性化学习路径规划研究[J]. 电化教育研究,2019,40 (10):5360.
- [31] 王雯霞. 图书情报硕士研究生群体的用户画像研究[J]. 高校图书馆工作,2022,42(2):2833.
- [32] REN C L, WANG Xuelong. Research on students' performance portrait method based on multidimensional data[J]. Journal of physics:conference series,2019,1237(3):0320436.
- [33] 彭红超,祝智庭. 人机协同决策支持的个性化适性学习策略探析[J]. 电化教育研究,2019,40(2):1220.
- [34] MULDER S, YAAR Z. The user is always right: a practical guide to creating and using personas for the web[J]. Information research an international electronic journal,2006,55(1):7476.
- [35] 孙小然,陈谊,斗海峰. 高校本科生毕业去向与课程成绩的关联可视分析[J]. 中国科技论文,2023,18(3):330-339,350.
- [36] 李澄锋,陈洪捷,沈文钦. 中外联合培养经历对博士生科研能力增值及论文产出的影响:基于“全国博士毕业生离校调查”数据的分析[J]. 高等教育研究,2020,41(1):5867.
- [37] 钱嫦萍,徐玉兰. 延期博士生的“佛系”心态:生成、表现及诊治路径[J]. 学位与研究生教育,2021(1):7883.
- [38] 郭海燕,刘春荣,张志斌,等. 博士学位论文评审制度改革及其成效分析[J]. 中国高等教育,2019(5):2831.
- [39] 王倩,蒋贵友. 博士生学术志趣缘何消退:基于2022年*Nature*全球博士生调查数据的分析[J]. 高校教育管理, 2023,17(4):102113.
- [40] 别敦荣. 谁有资格“读博”:基于个人经历的认知[J]. 学位与研究生教育,2023(12):+11.
- [41] 刘贤伟,马永红. 社会资本对校所联培博士生创新能力的影响研究:基于心理资本的中介作用[J]. 高等工程教育研究,2017(5):175180.
- [42] 李澄锋. 收紧培养关口能否降低博士生延期毕业率:基于“全国博士毕业生调查”数据的分析[J]. 高等教育研究,2023,44(6):6778.

- [43] 赵金敏,周文辉,付萌萌. 雄心与焦虑:学术志趣何以影响博士生焦虑心理[J]. 研究生教育研究,2023(03):24-29,62.
- [44] WOOLSTON C, O' MEARA S. China's PhD students give their reasons for misery[J]. Nature, 2019(575):714-713.

(责任编辑:吴朝平 校对:杨慷慨)

An Exploration of Association Rules for Doctoral Graduation Factors: Group Image Features and Type Differences

WANG Ting¹, GENG Youquan², HUANG Yaji¹

(1. Graduate School, Southeast University, Nanjing 210096, China;

2. Research Center for Learning Science, Southeast University, Nanjing 210096, China)

Abstract: Doctoral education aims at cultivating high-level talents with academic innovation ability, shouldering the important task of knowledge production and promoting economic and social development. In the context of big data, the educational data of doctoral training process was used to explore the association rules between doctoral graduation time and training links, discover the main reasons for delayed graduation of doctoral students, and analyze the group characteristics of doctoral graduates at different times, providing objective scientific guidance for strengthening the management of doctoral training and improving the rate of doctoral graduation in universities. Based on the acquisition of five types of data from the management big data system of graduate students in a university, namely student enrollment information, basic information of supervisors, comprehensive quality performance, campus behavior data, and research and learning situation, data processing was integrated to form a data set, feature extraction and label system construction were carried out, and FP-Growth algorithm was used to explore the association rules of all labels and doctoral graduation factors. Finally, the group portraits of three kinds of doctoral students who graduated on time, graduated after the deadline and failed to graduate after the deadline were depicted, and the different characteristics of the three kinds of group portraits were compared and analyzed. The results show that there is a strong correlation between the admission method, course score, proposal situation, the number and completion time of scientific research results, whether they won scholarships and honors during study, and the graduation time of graduation thesis. The age of doctoral students, whether the university is "double first-class" before entering the university, the age of the supervisor, the years of the supervisor as a doctoral supervisor, the number of students under the guidance of the supervisor, the campus behavior and the graduation time of doctoral students are not strongly correlated; there are two kinds of typical group portraits for scheduled and delayed doctoral students, and three kinds of typical portraits for delayed doctoral students. The differences of the group portraits of three kinds of doctoral students mainly lie in three aspects: academic interest, academic foundation and academic ability. In order to improve the quality of doctoral training and improve the graduation rate of university doctoral students, it is necessary to break the rigid restrictions and invisible discrimination to ensure the effectiveness of the selection of excellent students; the second is to strengthen the process supervision and shunt out to ensure the scientific management of the training process; the third is to pay attention to curriculum supply and basic training to ensure the fluency of scientific research practice; and the fourth is to protect academic aspirations and physical and mental health of doctoral students, and to motivate the enthusiasm of scientific research, innovation and exploration.

Key words: doctoral education; portrait of doctoral students; graduated on schedule; delayed graduation; overdue graduation