

一、标题:基于大数据分析的研究生质量监测及信息服务研究¹

二、课题组情况

课题完成单位:沈阳建筑大学

课题负责人:孙焕良

课题组组主要成员:刘俊岭,片锦香,孙丽梅,许景科,温全,刘天波,牛志成,王守金,王永会,袁帅,郭耸,师金钢。

三、摘要

大数据时代为研究生质量评价与监测提供了一个新的机遇,本课题利用大数据技术,实现对研究生培养过程数据、研究生在校活动数据、研究生就业等各类数据进行处理与分析。研究基于大数据分析的研究生质量监测与信息服务体系、大数据获取与处理方法、研究生质量信息服务方法等。实现了研究生整个培养周期的更细粒度的质量监测,为研究生管理者、导师、研究生等多类型用户提供动态的监测预警服务,提升了研究生质量监测的水平。

关键词: 大数据; 行为轨迹数据; 研究生质量监测; 大数据分析算法

四、主要内容

我国研究生教育已经进入了质量时代,加强质量保证和监督体系建设,在学位与研究生教育事业发展中具有重要作用。

不断扩大的招生规模及丰富的教育培养类型增加了研究生质量评价与监测的难度。在研究生招生规模方面,仅2015年招生人数达57.4万人,每年的招生人数及在读研究生总数均位居世界前列;研究生的教育培养类型也不断丰富,基本形成学术型与应用型人才并重的培养格局。同时,研究生的教育质量评价的多维度性使得质量监测更加复杂,体现在评价研究生质量时需要综合考虑研究生培养过程中的多个质量指标,包括教学活动、教学师资、学科建设、学术成果、科研项目、创新基地、学术平台及杰出校友等。因此,研究生质量监测已成为一项具有挑战性的工作。

大数据时代为研究生质量评价与监测提供了一个新的机遇。大数据是数据分析的前沿技术,其可以从各种各样类型的数据中快速获得有价值信息。通过大数据技术可以实现对研究生培养过程数据、研究生在校活动数据、研究生就业等各类数据进行有效处理。同时,利用大数据分析技术,可以实现对海量、多源异构的研究生质量数据进行监测、分析和管理的,从而提升研究生质量监测的水平。

本课题利用大数据技术,实现对研究生培养过程数据、研究生在校活动数据、研究生就业等各类数据进行处理与分析。研究基于大数据分析的研究生质量监测与信息服务体系、大数据获取与处理方法、研究生质量信息服务方法等。实现了研究生整个培养周期的更细粒度的质量监测,为研究生管理者、导师、研究生等多类型用户提供动态的监测预警服务,提升了研究生质量监测的水平。

具体取得的研究成果包括:

研究成果之一:基于大数据分析的研究生质量监测与信息服务体系

本部分研究了基于大数据分析的研究生质量监测与信息服务体系组成及各部分之间的关系,建立了研究生质量监测与信息服务体系模型如图1所示。图中部分1为研究生质量相关大数据,这些数据涵盖了研究生培养过程中及毕业后的数据;部分2将研究生大数据处理为可用于监测分析的数据;部分3为与研究生质量相关的评价指标;部分4为大数据分析算法;部分5通过分析数据与评

¹ 基金项目:中国学位与研究生教育学会研究课题“基于大数据分析的研究生质量监测及信息服务研究”(编号:2015Y0414)。

价的相关性，选取与评价任务相关的数据，采用大数据分析方法，实现研究生质量的评价、监测与预测；部分 6 实现多类型用户的信息服务。其中，信息服务部分与质量监测、分析算法及大数据部分均有反馈，根据信息服务需求实现动态调整所需的数据及分析算法。

基于大数据分析的研究生质量监测与信息服务体系发表于成果 1。

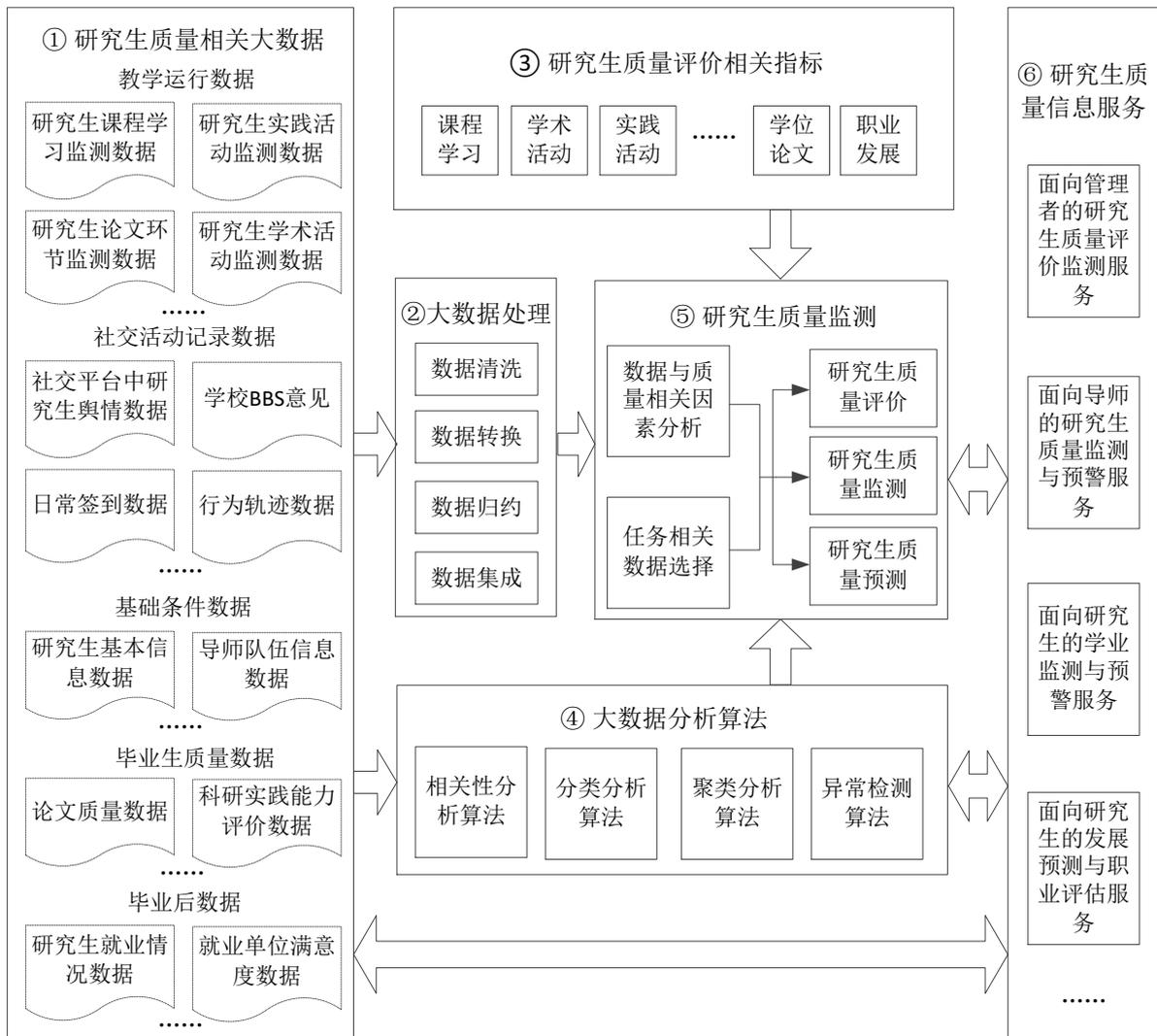


图 1 基于大数据分析的研究生质量监测及信息服务模型

研究成果之二：研究生质量相关大数据获取与处理方法

研究生质量相关大数据主要包括教学运行数据、社交活动记录数据、基础条件数据、毕业生质量数据等。现有评价研究生质量主要通过评估研究生规模、研究生就业、导师队伍、专业学位人才培养、国际化和满意度等维度。然而，在大数据背景下，许多与研究生学习生活就业相关的细粒度数据都可以获得，例如研究生的社交活动数据可以从社交媒体获取；研究生的实践活动情况可以从培养单位和社会组织多方面获取；社会评价可以从用人单位得到反馈。这些多渠道、异构的信息汇集到大数据分析系统中后，借助大数据分析方法，可以输出更细粒度的研究生质量监测数据。同时，需要研究研究生质量相关大数据的处理方法，去除数据中的噪声、不一致。

采用大数据处理技术，对研究生质量大数据进行预处理，去除数据中的噪声、不一致，并对数据进行集成、归约、变换处理，为大数据分析算法提供高质量的数据进行分析。

研究生质量相关大数据获取与处理方法发表于成果 2、7。

研究成果之三：研究生大数据与研究生质量评价指标体系和监测方法

分析海量、多源异构的研究生大数据，结合研究生质量的相关评价指标，研究大数据与研究生质量指标之间的关联关系。研究生培养过程的质量监测所涉及的质量指标包括研究生的课程学习、学术活动、实践活动及学位论文等质量指标。在评价研究生质量方面，通过研究我国学位与研究生教育质量国家标准体系中的国家标准、学科/行业/领域标准、培养单位的质量标准等三个层次标准，确定了本课题的质量评价指标。所研究的研究生质量监测侧重于具体的培养过程以及相关的质量保障因素。

本部分研究基于大数据分析的研究生质量评价方法、基于多层次、细粒度的大数据的研究生质量监测方法、大数据分析算法在研究生质量监测的应用方法。在大数据时代，可以获取更细粒度的研究生数据，借助大数据分析工具中的相关性分析方法、分类分析算法、聚类分析算法、异常检测算法等，可以得到更加详细的研究生质量评价结果。动态细致地评估研究生在学业、实践、科研、就业等方面的能力，为多类型用户提供监测及预测服务。

本部分内容发表于成果 2-3。

研究成果之四：基于研究生大数据的学业异常学生预测

本部分利用研究生活动数据分析学生的积极性、学习的投入时间、生活规律性等，从而用于预测学生的成绩，重点发现有学业异常的学生。研究内容包括数据预处理、属性相关性分析、预测算法等。

预处理部分实现数据的清理、集成，将细粒度的、多类型的数据处理成适用于应用的形式。对影响学业成绩的主要因素进行分析，计算其与成绩的相关性系数选取早饭次数、图书馆门禁次数等属性。其次，利用基于线性回归的 Lasso 模型对属性进行筛选，找到显著与学业成绩相关的属性。然后，采用两类方法进行异常学生预测，一类是基于分类的算法，采用 C4.5、CART、朴素贝叶斯进行异常学生预测；另一类是预测算法，采用 Lasso 算法进行分数预测，然后根据数据分布情况确定异常。实验结果表明 Lasso 预测模型适用于预测成绩后确定异常，对于学业异常学生的预测稳定性更好。分类算法适用于预测前对异常学生有明确标注的情况，能够直接预测出学生的类别，对于学业异常学生的预测准确率更高。

采用真实的校园一卡通数据进行实验，验证了学生的成绩与其日常行为规律有着密不可分的关系。预测模型能够广泛应用于学业异常学生的预测中。

本部分内容见发表成果 7。

研究成果之五：基于研究生大数据的学生成绩预测

本部分研究利用研究生活动数据预测学生成绩。现有的研究方法利用学生的入学成绩、学生管理老师的记录、授课教师的课堂记录，由于学生活动规律的变化性，这种方法预测的粒度为一个学期，即难以实时动态监控学生的行为，预测的时效性难以保证。

本课题利用学生活动轨迹数据，结合学生心理动机等因素实现动态预测学生的成绩。由于研究生大数据存在问题不完全、不充分、含噪声、异构、粒度过细等问题。同时，影响成绩的因素也较多，如学生的作息時間、同学关系等方面，如何选取属性、如何选取数据、采用什么方法进行预测是本研究的一个挑战。本课题借鉴心理学理论，发现一些相对稳定因素，通过一些导出的数据进行预测。心理学理论可知：学生的活动受内在动因驱动，内在因素不可见，力量强大。本研究将活动数据对学生的心理成就动机等属性进行标注，然后采用标注后的属性进行预测，较大程度上提出了预测的精度。利用真实数据集进行了实验验证预测分数在 5 分以下。

本部分内容见发表成果 8。

研究成果之六：研究生质量的多类别用户信息服务软件系统

通过搭建多类别用户信息服务平台，集成各种研究生相关质量数据，使得研究生教育管理者、研究生导师、研究生共享同一数据资源。根据不同类别用户的实际需求，形成有针对性的研究生质量信息服务产品，提供全方位、多层次、高效率、动态的信息服务。本部分研究所提出的信息服务的实践方法，结合所在学校的研究生教育，设计相应的软件，提供针对研究生管理者、导师及研究生的信息服务。

根据研究生教育管理者、导师、研究生等用户的需求，建立面向不同类型用户的质量监测信息服务系统。研究质量监测信息服务软件的实现方法，设计实现相关软件系统。根据研究生质量监测和预测结果，为决策者提供决策支持，为导师提供指导方向，为学生提供学习和就业指导。

本部分内容见发表成果 4-6。

本课题共发表学术论文 3 篇，硕士学位论文 2 篇，获软件著作权 3 项。发表成果如下：

[1] 王守金, 李婷, 孙焕良。基于大数据的研究生质量评价实施研究。教育, 2016:11 (9), 198.

[2] 片锦香, 浦春钰, 孙焕良, 侯静。大数据时代下的研究生教育质量评价体系研究。教育教学论坛, 2016, 6 (27): 75-76。

[3] 片锦香, 孙焕良, 张锐, 栾方军, 刘美佳。研究生教育质量评价体系趋势研究。中国校外教育, 2016, 5:36, 54。

[4] 孙焕良等。基于行为轨迹数据的异常对象检测系统 V1.0 (2015SR209698)。计算机软件著作权, 2015.10.

[5] 孙焕良等。结合空间位置和社会关系的服务类任务分配系统 (2015SR209702)。计算机软件著作权, 2015-10-30。

[6] 孙焕良等。基于 B/S 的区域流动模式构建及预测系统 V1.0 (2015SR208561)。计算机软件著作权, 2015.10.

[7] 韩松阳, 孙焕良 (指导教师)。基于活动轨迹的学业异常学生预测研究, 沈阳建筑大学硕士学位论文, 2016.12。

[8] Olaoluwa ESHO, 孙焕良 (指导教师)。Predicting Academic Performance of College Students from Campus Big Data through Psychology, 沈阳建筑大学硕士学位论文, 2016.12。

五、研究结论及政策建议

随着高校信息化进程不断加强，与研究生质量相关的各种数据的获取成为可能。加之数据挖掘技术的迅速发展为从多源异构、动态、海量的大数据中挖掘出有价值的信息提供了技术保障。从大数据中获取关于研究生质量的相关信息是未来研究生质量评价体系发展的趋势。本课题的创新性在于充分利用和研究生相关的大数据，采用数据挖掘技术，建立大数据时代特征的研究生质量监测系统。结合定性和定量分析方法，挖掘数据背后隐藏的知识，建立具有辅助决策的信息服务平台，使研究生质量评价体系能够动态、及时、信息全面、多角度地对研究生质量进行质量监测和预警。同时，根据不断更新的大数据，能够动态调整研究生质量评价体系，使得研究生质量评价体系随着时代发展变化而发展。

课题中所研究的内容均实现了任务书的目标，其中，“基于大数据分析的研究生质量监测与信息服务体系模型”、“研究生大数据获取与处理方法”发表了学术论文，“基于研究生大数据的学业异常学生预测”与“基于研究生大数据的学生成绩预测”方面形成硕士学位论文，研究生质量的多

类别用户信息服务软件系统注册 3 项计算机软件著作权,相关系统已在线应用。以上研究实现了研究生整个培养周期的更细粒度的质量监测,为研究生管理者、导师、研究生等用户提供了动态的监测预警服务,切实提升了研究生质量监测的水平。

研究生质量监测和预警研究生教育的重要内容,特别是大数据技术提供了相应的技术支持,本课题的后续研究将深入学生的心理健康等方面的预测与分析,其意义更大深远。

建议在基于大数据的研究生质量监测方面加大支持力度。

六、参考文献

- [1] 赵必华. 教育评价范式: 变革与冲突[J].比较教育研究, 2003, 10(161): 62-66.
- [2] 吴宏杰, 聂文强, 余苑等. 浅论中国研究生培养[J], 电子制作;2013(13): 152-152.
- [3] 肖海涛. 论高等教育学制改革与教育现代化[J]. 高等教育研究, 2007, (2): 39-44.
- [4] 研究生教育质量报告编研组. 中国研究生教育质量年度报告(2014), 中国科学技术出版社, 2015,北京.
- [5] 中国学位与研究生教育发展年度报告课题组.中国学位与研究生教育发展年度报告 2013, 中国人民大学出版社, 2014, 北京.
- [6] 王战军, 李明磊. 研究生质量评估: 模型与框架[J], 高等教育研究, 2012.33(3): 54-58.
- [7] 雷彦兴, 王德林. 美国当代学位制度的特征[J], 学位与研究生教育, 2002,9: 38-41.
- [8] 郑金娥, 祝湘陵. 中国与美国学位制度比较初探[J],长春教育学院学报;2010, 26(2):56-57.
- [9] B. Charles Tatum, Peggy Hayward, Rey Monzon. Faculty Background, Involvement, and Knowledge of Student Transfer at an Urban Community College [J]. Community College Journal of Research and Practice, 2006, 30: 195-212.
- [10] 袁锐锷. 中英教育硕士专业学位教育的比较研究[J], 比较教育研究, 2000(3):26-29.
- [11] 夏人青. 欧美国家高校与政府关系的比较研究[J], 北华大学学报(社会科学版);2003,4(4):57-61.
- [12] 李国杰,程学旗. 大数据研究:未来科技及经济社会发展的重大战略领域—大数据的研究现状与科学思考[J]. 中国科学院院刊. 2012,27(6):647-657
- [13] 教育部、国家发展改革委、财政部. 关于深化研究生教育的意见, 教研[2013]1 号, 2013.3.
- [14] 孙崇文. 国际研究生教育发展的趋势及其现实启迪[J]. 高等师范教育研究, 1997, (2): 38-46.
- [15] 姚利民,王燕妮.课程教学培养研究生科研能力之对策[J], 黑龙江高教研究;2006,9(149):89-91.
- [16] 梅仪新. 高等教育制度比较研究热点问题回顾与思考[J], 教学研究; 2007, 30(5): 381-384.
- [17] 莫蕾钰, 洪成文. 教育研究如何跟上大数据时代, 光明日报, 2015.04.02.
- [18] A. Acharya and D. Sinha, "An Intelligent Web-Based System for Diagnosing Student Learning Problems Using Concept Maps," Journal of Educational Computing Research, May 2016.
- [19] Scott-Clayton, Judith. "Do High-Stakes Placement Exams Predict College Success? CCRC Working Paper No. 41." Community College Research Center, Columbia University (2012).
- [20] Roderick, Melissa, Vanessa Coca, and Jenny Nagaoka. "Potholes on the road to college high school effects in shaping urban students' participation in college application, four-year college enrollment, and college match." Sociology of Education 84.3 (2011): 178-211.
- [21] Robbins S B, Lauver K, Le H, et al. Do Psychosocial and Study Skill Factors Predict College Outcomes? A Meta-Analysis[J]. Psychological Bulletin, 2004, 130(2):261-88.