

课题名称：基于大数据的研究生质量评价
指标体系研究

课题委托方（甲方）：中国学位与研究生教育学会

课题依托单位（乙方）：天津大学

课题负责人：白海力

课题组主要成员：孙鹤 王玮 余辉 王梅 米靖 杨院

起止日期：2015年7月至2017年3月

填表日期：2017年3月30日

一、摘要

随着高校研究生教育的不断发展，研究生培养质量评价指标体系的建立至关重要。本课题依托“大数据”技术，规划建立研究生培养质量评价指标体系；设计信息化云平台尝试对研究生培养过程进行全周期的动态跟踪和定量分析；以天津大学精仪学院生物医学工程专业20年研究生人才培养为案例样本开展案例库建设，设计BigData架构数据库并在自主搭建服务器集群云平台上进行部署，提供研究生培养质量信息服务，为研究生培养体系建设提供决策信息支持。

关键词：研究生教育；大数据；研究生质量评价指标体系；内涵式发展

二、主要研究成果

1. 面向人才培养的质量评价指标体系

通过对欧美各国研究生教育质量评估体系进行分析与探讨,借鉴其成功经验,对于建立符合我国国情的研究生教育质量评估体系,保障和提高研究生教育质量,促进我国研究生教育持续健康发展,具有重大现实意义。因此,项目组去芜存菁、博采众长制定了具有三级评价指标项的天津大学研究生培养质量指标体系,

研究生培养指标中的条件指标为研究生培养质量的形成提供入口保障,过程指标体现了研究生培养质量形成的过程质量,而成果指标则是研究生培养产出质量最重要的评价依据。因此,从研究生培养质量评价体系可以看出,它是条件指标、过程指标和结果指标的统一,逻辑结构上划分为5个一级评价指标,把5个一级评价指标划分为12个二级评价指标,然后把12个二级评价指标划分为30个三级评价指标,遵循了研究生培养质量评价体系设计的原则,为研究生培养质量的评价提供了一个全面而完整的三级评价指标体系。

2. 研究生培养质量量化评价模型

结合天津大学研究生培养现状,项目组制定了全过程培养质量量化评价模型,对每一个研究生在校期间培养质量进行全程动态量化考量。该模型兼顾研究生在校期间四方面的培养状况,即思想政治考核得分、学业成绩、科研成果得分、以及其它社会活动获奖与违纪处分加减分。

研究生培养质量量化得分= (思想政治考核得分×15%) + (学业成绩×40%) + (科研成果得分×30%) + (其它社会活动获奖与违纪处分加减分×15%)。

3. 以天津大学生物医学工程系研究生培养为例，进行案例化研究

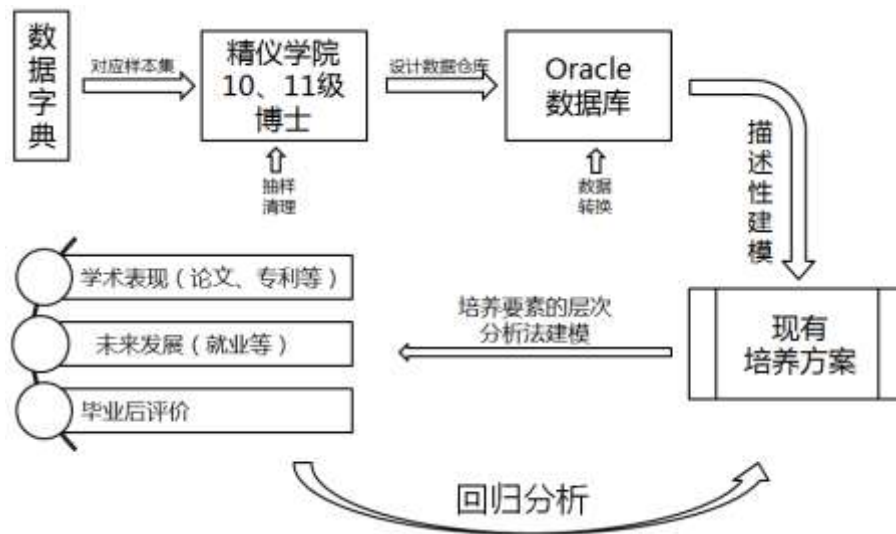


图1 生物医学工程系研究生培养案例化研究

为验证大数据的研究生质量评价指标体系，项目组以天津大学精仪学院生物医学工程专业20年研究生人才培养为案例样本，完成数据收集梳理，设计BigData架构数据库（研究生个人信息数据库）（图2）并在自主搭建服务器集群云平台进行部署，提供研究生培养质量信息服务。目前该平台已经申请国家软件著作权（图3），预计2017年4月底即可获得授权。此外该平台同时为全国生物医学工程专业学位研究生教育指导委员会主办的全国生物医学工程专业学位研究生教育网提供后台数据支持（图4）。

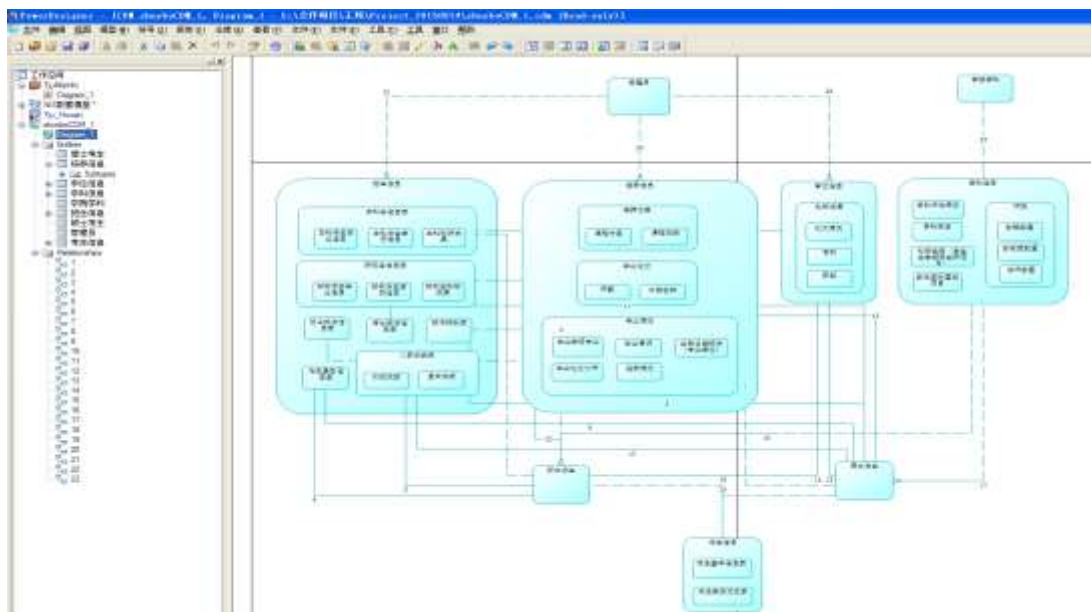


图2 大数据云平台生物医学工程系研究生培养数据库 CDM 模型



图3 国家软件著作权登记受理通知书



图4 大数据云平台案例化研究实例——生物医学工程系研究生培养

鉴于历史数据存在缺失等现象，实际案例研究中我们采用表 1 所示分级评价模型进行研究生培养质量的量化分析。

表 1 基于大数据研究生培养质量分级评价模型

一级指标	二级指标	三级指标	观测点	指标说明	分值	权重
入学评价	个人特征	学习成绩		本科（硕士）或研究生（博士）阶段课程成绩	6	16%
		获奖		奖学金、竞赛、科研、参与社会活动等获得的个人荣誉	4	11%
		科研成果	论文	按教育部学科评估体系评价	6	16%
			专利	按教育部学科评估体系评价	4	11%
		社会实践		参加学校组织、社会实践活动情况，担任学生干部情况，参加工作情况	3	8%
		个性与人格特质		基于心理测试的人群分类/个体报告的学情分析	基于适应程度的分析结果确定	
		入学考试表现	考试成绩	研究生入学考试成绩	8	22%
	综合评定		研究生入学面试成绩	6	16%	
	社会学变量	家庭背景		家庭成员、家庭环境、父母职业、父母学历、城乡等	4	19%
		专业背景		本科（硕士）或研究生（博士）阶段所学专业	6	29%
		学历背景	本科或专科	毕业院校水平	6	29%
			硕士	毕业院校水平（仅限博士）	5	24%
	培养评价		培养方案	课程体系质量	教学成果奖、精品课程	4

		学生成绩	培养计划成绩	6	22%	
		师资	代表性骨干教师能力评估	5	19%	
		国际交流	公派出国、国际会议	4	15%	
		支撑平台	国家重点实验室、2011 计划等	5	19%	
		科研项目	纵向、横向课题	4	15%	
		导师评价	导师指导能力、个性特征(基于与研究生个性特征匹配研究)	3	11%	
毕业评价		毕业质量	毕业论文送审及答辩情况	8	19%	
		就业情况	就业单位性质、层次, 岗位职责等	6	14%	
		科研成果	论文	按教育部学科评估体系评价	8	19%
			专利	按教育部学科评估体系评价	7	16%
		获奖	按教育部学科评估体系评价	4	9%	
		个性特质二次评价	基于心理测试的人群分类/个体报告的学情分析	基于适应程度的分析结果确定		
就业跟踪		职业发展	就业情况跟踪	5	12%	
		社会贡献	重大社会贡献	5	12%	

表 2 定量评价排名前 10 位博士统计信息

学号	综合得分	科研成果（单项排名）	毕业论文盲审	国家奖学金	海外经历
1012202068	96.5	20（1）	A+A+A	是	有
1013202023	95.2	18.5（2）	A+A+A	是	有
1010202023	93.4	15.5（4）	A+A+A	是	有
1012202069	93.4	16（3）	A+A+A	是	有
1012202071	92.3	12.5（6）	A+A+A	是	有
1012202061	92.1	13（5）	A+A+A	是	有
1011202069	91.5	12（7）	A+A+B	是	有
1012202067	91.1	11.5（9）	A+A+A	是	有
1012202065	90.7	11（12）	A+A+B	否	有
1012202064	90.4	11（12）	A+A+A	否	有

考虑到数据源的质量和本文的篇幅，仅对 2012-2016 年期间生物医学工程专业总共毕业 38 名博士为样本，按照表 1 评价模型进行定量评估，对最终排名和三级指标按照大数据挖掘理论中 apriori 算法，对境外交流经历与研究生培养质量进行关联性分析，对得到的 7 条显著性规则(致信度>0.85)进行阐述解释：

(1) 培养质量得分&境外交流经历强相关：培养质量综合得分排名前 15 位同学无一例外全部具备境外交流经历；

(2) 培养质量得分&奖学金强相关：培养质量综合得分排名前 10 位同学全部都获得过国家奖学金，20 名境外交流经历的同学都有获各类专项奖学金经历；

(3) 培养质量得分&科研成果&毕业质量高度相关：培养质量综合得分排名前 10 位同学科研成果得分单项排名都在 12 名以内（人均发表 SCI 论文 4.2 篇，授权专利 2.7 个），其中 8 人毕业论文盲审结果 3 个专家评价都是 A(质量最高档)；

(4) 培养质量得分&就业单位：培养质量综合得分排名前 10 位同学其中 5 位去高校就业，3 位去研究所就业，2 位去企业就业，但其余 28 名博士中有 19 位去了企业就业；

(5) 社会实践&就业情况&家庭背景：38 名样本中毕业后有 8 名从事行政管理工作，其中 7 位在校期间有过担任学生干部的经历，且全部 8 名同学家庭背景中父母中至少有一方在事业单位工作；

(6) 学历背景：38 名博士中有 24 人来自 985 高校，12 人来自 211 高校；

(7) 个性特质：38 名博士中有 28 位自认为性格内向，7 位自认为比较内向，仅有 3 位自认为性格偏外向。



图 5 大数据云平台下研究生培养数据网络信息服务系统架构示意图

4. 海量研究生数据的统计信息的可视化展示

数据分析统计方法不断改进。一般质量评价工作常采用传统统计方法，通过普通计算机或手工方式来处理结构化数据，而大数据时代的数据可以是半结构、非结构数据，其统计分析以相关关系为基础展开，数据分析统计方法也随着现代计算机技术的发展而不断改进。在海量大数据平台上可以采用html5等技术以图、树等人机交互友好的界面形式展示数据分析结果，见图6。



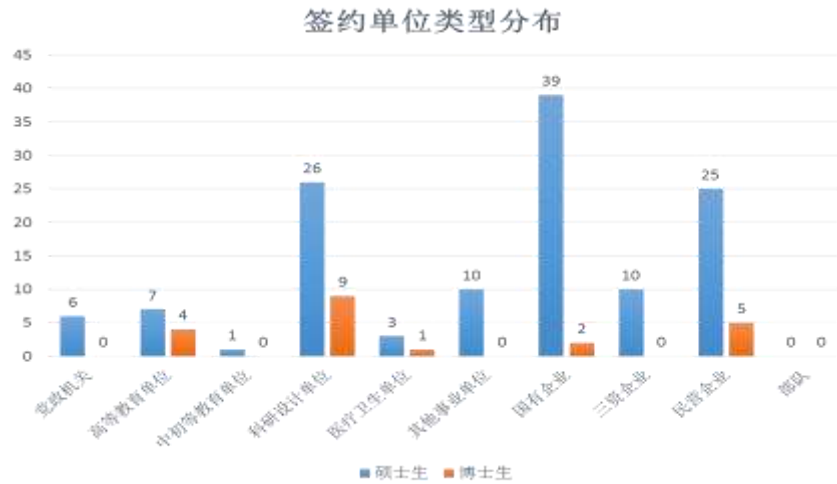


图6 海量数据基于多维数据集技术分布式跨库数据融合与检索

数据融合主要解决针对同一研究对象不同数据源之间的数据集成问题。例如：我们希望研究生物医学工程专业近 20 年历届毕业研究生的社会贡献做专题研究，便会遇到以下技术难题：

(1) 这部分数据时间跨度近 20 年，1984-1994 年间毕业生数据存于天津大学档案馆服务器，数据源为 excel 表格，1995-2012 年数据分布在学校招生数据库、研究生院综合信息数据库、精仪学院数据库；2013-至今数据可以从就业指导中心数据库、学院校友数据库及研究生院数据库中获取；

(2) 数据格式涉及 Excel、Txt 文本、Visual Fox、Oracle、微软 Access 及 SQL Server 等多个厂家、多个版本不同数据库；样本数据相互独立，数据库结构的复

杂性决定了来自不同服务器的样本数据并非由简单的叠加就能实现数据集成，因为数据库实体-关系模型本身要求保持各实体之间的关系完整性以及实体内部的唯一性约束。

多维数据集作为一种数据的再组织模式，在数据库和联机数据分析之间建立一中间层，在不影响数据源以及前台数据库程序的前提下，在维持数据库关系完整性和唯一性约束的条件下，通过对数据的再组织完成异种数据源数据的集成。多维数据集的数据共享主要是指共享维度的结构与信息。

三、研究成果及政策建议

大数据的特点决定了大数据下研究生培养质量评价工作存在的问题将逐步解决，同时，研究生培养质量评价工作的深度与广度必然将逐步发生相应的转变以适应形势发展。

1. 研究生培养过程的定量分析。研究生培养体系因其培养过程复杂、影响因素繁多而造成对其进行客观分析的方案可行性差，大数据技术的出现为这种数据量大、数据形式复杂多样的过程定量化分析提供了完美的解决方案。通过建立合理的定量分析模型、海量多样数据分析，可以实现对研究生培养各环节影响因素的量化呈现，为研究生教育决策提供可靠有效的数据支持。

2. 构建有效的研究生培养全链条质量评价体系。目前，评估体系多专注于单独的研究生培养环节，做一个有机整体，研究生培养全链条是一个闭环反馈系统，构建一个多维闭环评价体系才能更清晰的描述整个研究生培养体系的运作机制。通过对从招生到学位授予整个培养链条及相关学科建设等环节的综合考量，我们初步构建了一个多维度反馈的指标评价体系描述研究生培养过程。

3. 整合数据源。根据天津大学研究生院现有IT信息资源，研究生信息管理系统、教学评价系统，数据指标、招生、培养、学位授予、科研成果评定评价等基础数据涉及Access、Oracle、微软SQL等多种数据库，部分数据以Excel、Word文本等格式存储，项目组采用微软SQLServer2008中Analysis Server构建多维数据集，采用数据仓库技术进行跨平台、多种异构数据源集成。

4. 构建数据库。采用Sybase公司主流数据仓库工具PowerDesign，以公共仓库元模型为标准，基于上述架构为每个研究生建立个人信息档案，其中包括研究生本科（研究生）期间学业/学术表现、研招成绩综合评价、培养方案执行情况、学术表现（科研成果）、学位论文情况、导师评价、个人素质综合评定等方面以及研究生、导师问卷调查信息，设计数据字典，构建CDM模型。

5. 数据的可视化方案。采用html5等技术以图、树等人机交互友好的界面形

式展示数据分析结果。

6. 以分析结果为主体，辅以研究生、教师问卷调查分析结果，为学校政策的制定提供决策支持。

参考文献

- [1]Gantz J, Reinsel D. Extracting value from chaos[J]. Idcemc2 Report, 2011.
- [2]Mohammad S. Owlia. Quality in higher education-a survey[J]. Total Quality Management, 1996, 7(2):161-172.
- [3]赵立莹, 马亚菲, 田养利. 美国 NRC 最新博士专业质量排名:发现、评价、判断[J]. 学位与研究生教育, 2011(9):61-64.
- [4]吕进, 李成炜. 德国 CHE 高校排名与中国教学评估指标对比的启示[J]. 浙江科技学院学报, 2010, 22(5):360-364.