

# 大数据背景下的应用统计专业硕士 人才培养模式研究

阮敬陈涛

**摘要:** 本文针对当前大数据背景下的数据分析人才需求的特征和现状,以培养高端应用型人才为目标,探讨将大数据分析技术和统计分析技术相结合的实践教学模式,以培养学生对大数据的分析能力,并对我国应用统计专业硕士的大数据分析人才培养模式进行了探讨。

**关键词:** 大数据 大数据分析 应用统计专业硕士 人才培养

**基金项目:** 本项目受到首都经济贸易大学 2014 年教学改革项目“应用统计专业硕士大数据分析人才培养模式创新与实践研究”的资助

DOI: 10.3969/j.issn.1674-537X.2015.08.011

## 一、引言

随着信息技术与统计学理论的深入发展,大数据已经深入到国民经济生活的各个方面,大数据分析日益成为人们社会经济生活进行预测和决策的重要手段,大数据分析能力也已成为现代社会高端人才应当具备的基本技能之一。

目前,大数据方面的研究在美国、日本等发达国家已经上升为国家战略。然而在我国高等教育人才培养中,对大

数据的关注还不够,传统数据收集和分析技术的知识结构已不能满足大数据时代的高要求,社会亟需具有大数据视野和分析能力的综合性人才。统计学专业作为研究数据的科学,已经在与传统计算机技术、数据库技术、网络技术等专业领域逐步融合,从学科发展趋势上看,需要对高层次统计专业人才培养,以适应大数据时代的来临。对此,我国已于 2011 年开始大规模招收应用统计专业硕士,培养数据分析领域的高级人才,但尚未有针对大数据分析高端人才的培养。因此,在调整和改革现有应用统计专业硕士人才培养方案的基础上,探索能够满足大数据分析需求的人才培养模式,设置合理的课程规划及其具体内容,使得统计类高端人才培养在大数据时代抓住先机,率先与国际接轨从而提升统计分析人才的竞争力具有重要意义。

## 二、大数据背景下数据分析人才的能力需求

大数据除去概念观其本质有二:第一,如何架构大数据环境来管理、存储和提取数据;第二,如何根据“在线”和“实时”等模式从数据中发掘出价值。

第一个本质其所关注和研究的是大数据硬、软件架构的问题,需要掌握服务器设备、存储设备以及必要组网设备(如

社会发展相对落后,当地居民的生活水平与其他地区的相比,也相对较低。因此,对于很多家庭经济条件较差的老人来说,经济上的压力会导致他们形成较严重的负面情绪,从而影响老年人的心理健康。

### (四) 社会文化

西昌作为凉山彝族自治州的州府,彝族人口占总人口很大一部分,其次,还有汉、藏、回等其他民族,各民族之间频繁地交流和融合。在这个过程中,不同民族间的文化冲突不可避免,由此带来的人格贬低与压力也影响着当地老年人,特别是彝族老年人的心理健康水平。每一个人都是本民族文化的活的载体,不同民族间的交流实质就是不同文化的交流。在交流的过程中,文化间的碰撞和冲突不可避免,而这所带来的是心理困境和文化适应问题。彝族老年人自小生活在民族聚居区,深受本民族文化传统的熏陶,在情感态度、思维方式、价值观念、审美情趣、生活习惯、行为方式上无不打上了本民族文化的深刻烙印。在与其他民族交往的过程中,他们对文化差异的体验尤为敏感深刻,在心理上引起的冲突也更加剧烈。同时,其他的民族有时甚至拿他们的宗教禁忌开玩笑,这让他们深深感觉不被理解、接受和尊重。这些使

他们情绪抑郁、人际交往敏感。他们迫切需要针对性地加强心理健康服务。要进一步提高民族地区老年人的心理健康水平,文化适应是不能回避的问题。

### 参考文献:

- [1] 王建英等. 郑州城区老年人心理健康状况及其相关因素分析 [J]. 中国老年学杂志, 2005(7).
- [2] 李娟. 北京城区老年人心理健康状况及其相关因素分析 [J]. 中国老年学杂志, 2002(9).
- [3] 解登峰等. 老年人心理健康状态的影响因素 [J]. 中国老年学杂志, 2011(1).
- [4] 解静, 陈元玉, 江琳, 王少冰. 汕头地区养老院和社区老年人生活质量、孤独感、自尊及影响因素的研究 [J]. 中国临床心理学杂志, 2011(03).
- [5] 胡月, 龚磊, 陈福宽, 孙大勤, 董昀球, 王莹莹, 康琦, 陈家应. 农村老年人自评健康状况的影响因素分析 [J]. 中国卫生统计, 2013(02).

(作者单位: 西昌学院经济与管理学院)

交换机)、大规模集群系统、基于跨物理服务器的多点系统数据库系统、分布式文件系统、MapReduce 计算框架、实时流计算、内存计算、图计算框架等软硬件综合知识;该方面是计算机科学领域所关注的重点技术和内容。

第二个本质所关注的内容主要为大数据应用以及不同数据的关联,可以主要细分为两个重要方面:1、如何将传统单机数据运算与数据分析方法应用到分布式文件系统的集群中以提升运算效率,需要进行算法应用的研究;2、如何在大数据架构基础上将已有分布式计算方法(统计分析方法、数据方法以及机器学习算法等)应用到解决实际问题当中。这两个方面是统计学关注的重点内容。

综合以上分析,大数据系统架构是计算机科学领域中较底层的分布式文件存储和管理系统(当前较为主流的开源体系为 Hadoop/MapReduce 框架,此外还有实时流计算、分布式内容计算、图计算框架等与之并存),大部分的商业软件如 SAS、Matlab 均可以将其直接整合,架构在其上(用户无需过多关注数据底层),面向用户提供数据管理和分析功能;同时一些开源软件如 R 和 Python、Perl、Ruby 等面向对象的开发性质脚本语言也可以运行于这些系统中,提供数据分析功能。

从美国 20 多所院校的著名大数据项目来看,他们的项目课程也是围绕上述两个本质性的问题来设计的。首先是以卡内基梅隆为代表的以大数据底层系统为主的项目,其所研究的是大数据系统架构的问题;其次是以北卡州立、麻省理工 MBA 等为主的大数据分析应用硕士,其所关注的重点内容为大数据在各领域中的应用问题。

### 三、大数据背景下应用统计专业硕士人才的培养与实践

从我国应用型人才培养的模式来看,应用统计专业硕士非常适合培养与国际接轨的大数据分析高端人才。因此,本文将结合笔者所在单位的人才培养经验探讨大数据背景下应用统计专业硕士高端应用型人才培养的模式和实践经验。

首先,对现有大数据分析理论进行梳理,搭建大数据理论与统计理论之间的桥梁,为应用统计专业大数据分析硕士人才培养提供学科基础。

大数据分析起源于互联网数据的高度聚集和流动。聚集数据往往以分布式形式存储在网络中的不同节点上,这需要对传统数据分析方法进行分解(MAP)和规约(REDUCE)上的改进才能快速得到分析结论;流动数据则大都以数据流的形式临时存储与内存当中,这同样需要对传统数据分析进行流分析的改进。培养单位要研究这些新数据分析方法与传统统计数据分析理论之间的区别与联系,有利于培养学生的大数据分析思维,从而为应用统计专业大数据分析硕士的培

养提供理论基础。

其次,充分掌握大数据分析人才的需求,以应用角度来设置应用统计专业大数据分析硕士的课程体系与内容,培养学生的大数据分析能力。

大数据分析方向的硕士课程应当着力于“分析”能力的培养,这样才能在计算机领域中突出统计特色,在大数据时代才能占有一席之地。即:将大数据架构基础上的数据存储和分析环境作为解决问题的基本工具,培养学生搜集数据、整理数据、分析数据、呈现数据的基本能力,并将这些能力应用于解决实际工作的具体问题。在培养分析能力的基础上,需要开设一些必备的分布式运算以及系统开发等计算机技能选修课程,从而满足社会广泛需求。

其三,将大数据分析人才培养目标与行业应用相结合,培养真正对社会有用的复合应用型人才。

行业大数据应用往往建立在已有大数据存储系统中,如主流的 Hadoop、Storm 等大数据分析环境。教学单位应将这些主流分析环境引入课程,在课堂中构建真实的数据收集、数据处理和数据分析环境,以案例教学的形式生动而真实的模拟实际解决问题的需求,让学生在校园中就能充分接触到行业应用的前沿,这有助于实现教、学、产、研真正相长。

其四,将大数据分析硕士的师资力量整合,实现科研部门、教学部门与实际行业部门之间的联合培养协同创新。

大数据时代的数据之间具备一定的必然联系,培养大数据分析人才的师资力量在不同学校、不同学院以及实际行业企业或部门中都是有联系的。应用统计专业硕士的大数据分析人才培养应将涉及到的计算机科学、信息科学、统计科学等及其他数据应用领域的实际部门中的师资力量进行整合,统一规划培养课程和具体内容,共同培养适应社会发展潮流的大数据分析高端人才。

其五,在现有应用统计专业硕士研究生培养计划中做好大数据课程的衔接工作并解决课程内容规划等具体问题,为大数据分析硕士培养提供制度保证和具体规划措施。

通过以上培养时间,以应用统计专业硕士为依托培养大数据分析人才,有助于整合高等院校、科研机构 and 行业实际部门的师资力量,以及校内外相关实习、实践、实训资源,以点到面充分发挥统计专业类学生的技术优势,通过紧密交流大幅提升教师的教学水平。同时,也拓展了现有应用统计专业硕士的研究方向,并与国际主流应用统计专业硕士接轨,能够使统计专业高端应用型人才培养走在全国前列并提升人才竞争力,在大数据背景下拓展本专业的就业方向 and 途径。

(作者单位:首都经济贸易大学统计学院)