

# 新工科背景下大数据专业导论课程的改革与探索

张祖平

中南大学计算机学院, 湖南 长沙 410083

## 摘要

在申报与建设数据科学与大数据技术专业的热潮中,专业培养体系与相关课程大纲一直是各个高校体现其特点的内容。针对专业导论课程的培养目标,结合新工科研究与实践项目的要求,论述了专业导论的教材准备、课程定位及具体教学内容。通过在专业课程中引入实践教学,实现专业能力体系的培养目标,让学生在进入大数据专业的初期就形成良好的实践意识,并对典型技术有切身体会,从而达到大数据专业的知识水平与能力要求。

## 关键词

新工科;大数据专业;导论课程;能力体系

中图分类号:TP392

文献标识码:A

doi: 10.11959/j.issn.2096-0271.2018059

## *Reform and exploration for introduction course of big data professional under the background of emerging engineering education*

**ZHANG Zuping**

School of Computer, Central South University, Changsha 410083, China

## *Abstract*

In the upsurge of declaration and construction about data science and big data technology professional, the professional training system and related curriculum syllabus have always been the characteristics of various colleges and universities. Aiming at the training objective of introduction to specialty course and the requirements of emerging engineering education, the teaching materials, course orientation and specific teaching contents of introduction to specialty were discussed. By introducing practical teaching into introduction to specialty, it reflects the training objective of professional ability system. Its aim is to form students a good sense of practice at the early stage and have a personal experience with typical technology. The knowledge level and ability requirement of big data professional will be achieved easily after introducing practical teaching into introduction to specialty.

## *Key words*

emerging engineering education, big data professional, introduction course, ability system

## 1 引言

随着移动互联网的兴起,全球数据呈爆炸性增长,数据规模大约每两年翻一番。而随着人工智能环境下物联网生态圈的形成,数据的采集、存储、分析处理、融合共享等技术需求都能得到响应,各行各业都在体验大数据带来的革命,与大数据相关的技术人才需求激增。据预测,中国近年内大数据人才缺口达200万人,全世界相关人才缺口超过千万之多,因此培养大数据技术人才的重任落到了高等院校身上。从事数据统计、分析和应用的人才已经难以适应大数据时代的新要求。大数据具有体量巨大、速度极快、类型众多、价值巨大的特点,对数据从产生、分析到利用提出了前所未有的新要求。高等教育只有转变观念,更新方法和手段,寻求变革与突破,才能在大数据与人工智能的信息大潮面前立于不败之地。我国教育部门为了响应社会发展需要,于2016年开始正式开设“数据科学与大数据技术”本科专业与“大数据技术与应用”专科专业。近几年,全国形成了申报与建设大数据相关专业的热潮。

随着大数据专业建设的推进,国内对专业知识体系与能力体系进行了探索<sup>[1-2]</sup>,同时也出现了对专业课程建设与人才培养模式的讨论<sup>[3-4]</sup>,力图探索大数据教育体系中可采用的具体策略和方法。从2017年开始,全国范围内开始讨论新工科,由此催生了《教育部办公厅关于推荐新工科研究与实践项目的通知》的发布,同时也开始了对新工科背景下大数据专业建设的探究<sup>[5-6]</sup>。大数据专业的培养体系在专业教育人才的通力协作下逐步形成,各类适用于不同高校大数据专业特点的培养方案与课程大

纲也日渐成熟。

作为专业重要课程之一的专业导论课程历来受到各类高校的高度重视,专业导论课程一般是在学生第一年统一基础教学后开展的,旨在给学生普及相关专业知识,引导学生进行专业学习。而在新工科背景下,大数据的专业导论具有不同的要求与特点。

## 2 专业导论课程大纲

“数据科学与大数据技术导论”是一门面向本专业的导论性课程,旨在让学生在大学入学最初阶段就对本专业的发展历史、知识结构、培养目标及要求及与数据科学与大数据技术相关的基础知识、典型技术、具体应用等有直观的认识。区别于新生课程的普识性介绍,该课程的相关内容偏专业,目标是让学生对本专业的知识及培养要求有相对全面而直观的了解,同时该课程也会概述性地介绍与计算机学科相关的内容及典型人物,以激发学生的学习兴趣,进一步了解设置新专业的历史背景与总体要求。

“数据科学与大数据技术导论”课程的基本要求包括以下几个方面。

- 知识方面:较好地掌握数据科学与大数据技术的发展历史及相关典型概念,如与数据相关的基本概念、与数据特征相关的测度概念及与大数据相关的5V特性等;了解典型的大数据分析环境包括的技术体系,如Hadoop;了解计算机典型的基础概念,如数据、算法;了解专业需要掌握的知识体系及课程要求;对大数据技术的典型应用有相对直接的了解,并能联想到生活中的大数据技术应用场景。

- 能力方面:使学生对本专业的课程

体系有区别与选择的能力,对典型的大数据分析环境的技术体系有一定的判别与选择的能力,对应用系统是否要用到大数据平台有一定的判别能力,对整个专业的知识体系有一定的预判与认知。

● 素质方面:对数据科学与大数据技术专业的相关基础知识有相对全面的了解,逐渐形成采用数据分析的思维解决实际系统需求的意识;能够通过网络搜索平台找到大数据分析平台需要的典型开源性工具软件,尝试通过网上教学视频进行安装与调试,逐步形成直观认识与一定的学习、操练兴趣;通过课外导学的模式,从网上大量相关的实例中得到启发,从而提升自主学习和终身学习的意识,形成不断学习和适应发展的素质。

从以上分析可以看出,课程大纲响应了新工科突出能力与新技术的特点,从知识、能力与素质上提出了具体要求。

### 3 教材的准备与课程定位

为了适应“数据科学与大数据技术导论”课程的课程大纲,笔者组织相关教学团队,编写了《数据科学与大数据技术导论》教材<sup>[7]</sup>。区别于数据科学导论方面的教材<sup>[8-9]</sup>或大数据技术导论教材<sup>[10-11]</sup>,笔者单位的教材既包括数据科学与大数据技术专业的发展历程、专业知识要求与技能基本要求等,也包括有关数据科学的基本概念、数据挖掘的基本方法及大数据分析的主要技术等,对大数据分析的各流程中采用的关键技术及核心技术进行了梳理,对主要的大数据技术生态体系进行了介绍,最后基于实际项目,介绍了医疗大数据与智能城市交通大数据,既为学生提供了基本的数据科学与大数据相关知识,又介绍了实际应用的技术与高层次平台或项目申

报需要表达的与大数据相关的内容,同时通过20个自主实验强化学生的实践能力。

本课程首先介绍了数据科学与大数据技术专业的产生背景与发展历史、专业的特点与综合要求、专业相关的完整知识体系与技能体系,之后介绍了与本专业密切相关的专业,如计算机科学与技术、统计学等,分析了其与这样的专业的关联关系,还对数据科学与大数据技术专业的出路与就业情况进行了简述。专业课程体系总体架构如图1所示。

由于数据科学与大数据技术专业较新,可供借鉴的历史不多,大家各自的理解不同,在课程体系方面形成了“百花齐放”的局面,究竟哪个是最好的、最完整的,暂时没有结论。笔者提出的专业课程体系供大家参考,也供本专业学生选择课程时参考。

针对以上专业课程体系,专业导论课程需要将专业主要的知识点串接起来,既要各主要课程中抽出重点的知识进行综述性的讲解,又要兼顾各知识间的关联关系,同时由于课时的约束,内容不可能讲得很多,也不可能讲得很深,这是一个较难掌握的平衡度。

在技能体系方面,笔者认为数据科学与大数据技术专业的学生需要学习从数据获取到数据分析应用整个流程的各种技术与技能。目前,对本专业的学生或从事本专业技术工作的人员在专业技术方面没有明确的规定,但经过了几年的专业建设与讨论,并综合考虑了社会对专业人才的实际需求,形成了如图2所示的专业技能体系,不要求学生掌握所有的技术,但学生需要对某些部分或环节有选择的能力与熟知其中一二的基礎。

一般的专业导论课程主要是专业知识体系的串接,即使讲到技能体系,也只是一个概括性的介绍,如概述技术的名称与具

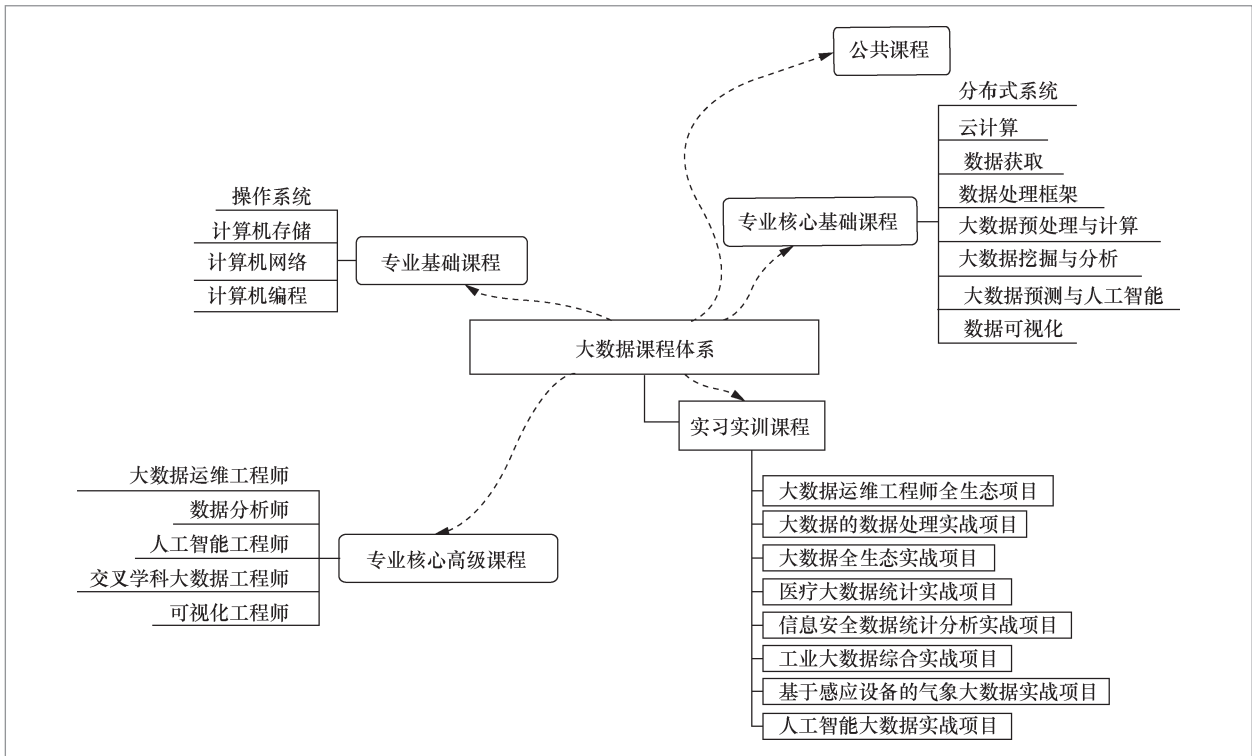


图1 专业课程体系总体架构

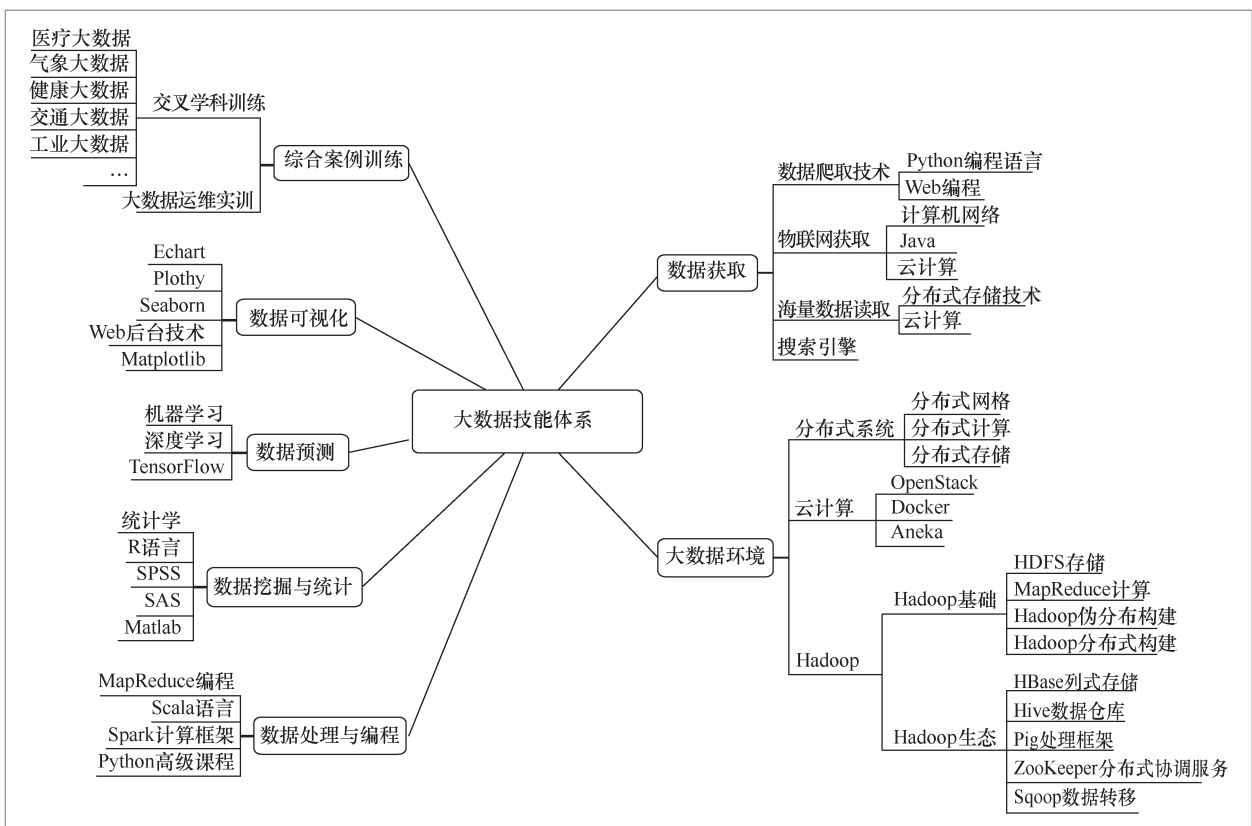


图2 大数据技能体系

体技术特点或要求等,不会要求学生有实践动手的机会。但从以上技能体系分析来看,数据科学与大数据技术专业要求掌握的技术可以说既具体又丰富,而且复杂性高。在新工科背景下,如果数据科学与大数据技术专业的学生在专业导论课程的学习时,没有形成自己动手的意识或基本的动手能力,就很难达到能力要求,也很难满足专业就业与科研的需要。因此笔者学校在实际教学时,专业导论课程对学生的动手能力提出了具体的要求。

## 4 专业导论课程教学内容

针对专业导论课程的大纲及定位,笔者学校将实际课程教学分为5个章节,其中第1章为专业概论,主要介绍数据科学与大数据技术专业的产生背景与发展历史、专业的特点与综合要求、专业相关的完整知识体系与技能体系,还介绍了本专业和与其密切相关的专业(如计算机科学与技术、统计学等)之间的关联关系,本章还对本专业的出路与就业情况进行了简述。

第2章为数据科学与大数据基本概念,主要介绍与数据科学与大数据技术相关的基本概念、相关技术特点、对应的社会岗位需求及对学生的知识、能力、素质要求。与数据科学与大数据技术相关的主要概念包括基本概念(如信号、数据、信息、知识等)和成体系的概念(如数据科学、数据挖掘、数据库、大数据等)。

第3章为大数据核心技术,大数据技术主要有6个核心部分:数据采集、数据存储与管理、数据预处理、数据清洗、数据挖掘、数据可视化。当然也有很多文献把大数据技术划分为5个部分,即将数据预处理合并到数据采集或数据清洗中。总体来说,大数据核心技术

是大数据处理的各个核心环节的关键技术。

第4章为大数据环境与技术,主要介绍大数据运行环境及典型技术。大数据技术生态一般是指在数据采集、数据整理、存储、运算、数据展示以及系统维护等各个层面用到的各类相互关联的技术、软件、工具等的集合。以Hadoop为例,其维护工具是Ambari,采集工具是ETL,管理工具是Sqoop、NiFi、Phoenix等,存储工具是HDFS、HBase、Hive等,运算工具是MapReduce、Spark等,联机分析处理(online analytical processing, OLAP)的关键工具是Kylin,数据展示包括很多技术或工具,Hadoop自带的是Zeeplin。

第5章为大数据应用系统,主要介绍2个典型的大数据应用系统及实用技术展示,包括医疗大数据、交通大数据等具体应用实例,进一步强化大数据相关技术。此外,部分相关项目立项时的一些文档内容可供大家将来参与类似项目时参考,如医疗大数据主要参考“医疗大数据应用技术国家工程实验室”“医学大数据协同创新中心”“数据科学与大数据技术专业”等的申报材料,交通大数据主要参考“大联合交管中心系统”的初步设计、详细设计、技术报告及用户手册等文档。

## 5 专业导论的实践内容

为了响应教育部新工科研究与实践项目的要求,从本专业导论课程开始设立相关的实验环节,通过知识主线与技术主线把相关课程串接起来,力争让学生尽早有培养自己动手能力的意识与综合利用各种技术与平台的能力。

首先是基本编程能力的强化训练,如

讲述第1章绪论时引出4个实验。

- 实验1: 任选编程语言, 实现自然数阶乘累加, 如 $1! + 2! + 3! + \dots + n!$ , 其中 $n$ 为输入变量。

- 实验2: R for Windows下载与安装、测试演示(graphics)、测试数学函数。

- 实验3: Matlab下载、安装、测试演示、测试 `fplot()` 函数并生成曲线图。

- 实验4: ECharts下载、安装与典型图表可视化。

大二学生已经有了一定的编程基础, 此时, 再强化训练细节编程(阶乘累加中, 当 $n$ 较大时需要细致处理)及常用统计分析软件的常用功能(如R语言、Matlab及可视化工具ECharts等)的使用。这有一定的挑战性, 但是图形结果的呈现也会给学生带来一定的兴趣。比如, 实验1看起来比较简单, 一般理解就是一个累乘循环、一个累加循环就可以了, 但实际上, 由于累乘的结果增长很快, 需要考虑整数的位长问题。当一般的位长不够时, 就需要考虑如何保存精确的整数, 当 $n$ 增长到比较大的数值(如20以上)时, 处理起来非常复杂(不同编程语言有不同的位长限制), 这就要求有较好的综合编程能力。有些学生可以做到结果用文件输出,  $n$ 只受计算能力的限制; 有些学生开始只能做到 $n$ 为10左右, 但经过多次尝试,  $n$ 就可以越来越大。经过这种从看起来简单到实际具有一定挑战性的综合编程训练, 学生对自己选择的编程语言的感受是非常深刻的。

在第2章安排了4个实验, 具体如下。

- 实验1: 任选编程语言, 实现数组的集中趋势测度。

- 实验2: 任选编程语言, 实现数组的离散程度测度。

- 实验3: 下载并安装Oracle, 实现在数据库管理员(database administrator,

DBA)用户系统下的用户管理。

- 实验4: 执行典型的SQL操作等。这部分的实验逐步转向专业, 即数据科学中的数据统计与数据库的相关操作, 难度不大, 但有专业性。

其中第2章中的实验1、实验2主要是数据统计方面的实验, 是一些计算公式的实验, 难度不是很大, 但能形成数据统计的感性认识。实验3、实验4主要是数据库操作, 目标是让学生对大型数据库的用户管理及基本的SQL中九大命令的操作有真实的感受。

在第3章安排了5个实验, 包括: 通过开放数据库链接(open database connectivity, ODBC)导入/导出数据; 网络蜘蛛的搜索与应用; SQL查重与去重; 利用SQL实现数据集成; 使用ECharts与Excel实现数据库表的数据可视化。此部分更趋向于专业, 且要求比较综合, 逐步提高了实验的难度与综合性。此章的实验逐步向大数据技术靠近, 有多种方式的数据采集、数据的预处理、数据整合与融合及分析结果呈现等, 各个实验都有一定的设计性与综合性, 要求也就相应地提高了。

第4章介绍了典型的大数据技术的相关操作, 而第5章安排了2个设计性的实验, 包括: 分类统计重症肌无力诊疗数据库中的首发症状类别及与年龄的关联关系; 基于交通大数据中环线路面卡口研判。难度不言而喻, 同时也不要要求所有学生都做到, 力争通过演示的形式, 给大家一个参考, 让学生感受真正的大数据分析与一般的统计分析差别到底在哪里, 从而使学生对大数据实际分析有一个感性的认识。

## 6 结束语

我国于2016年开设了数据科学与大

数据技术本科专业,国外虽然有数据分析工程的本科专业或硕士专业<sup>[12]</sup>,但专业培养方案与课程大纲可供借鉴性较小。本文针对数据科学与大数据技术专业中重要的专业导论课程,从知识体系与能力体系等方面分析了培养的要求与目标定位,在知识点的串接、动手能力意识的培养与实际操作方面提出了可供参考的方案,该方案的特点是针对新工科的要求,在具体教学中引入了大量的专业实践,为学生形成良好的能力训练意识与真实感受大数据提供了机会。未来需要进一步完善的是针对课程教学与教材,形成数据科学与大数据专业实践资源管理与服务平台。

## 参考文献:

- [1] 姚乐,朱启明. 赋能大数据教育: 全国高校大数据教育教学经验谈[M]. 北京: 电子工业出版社, 2018.  
YAO L, ZHU Q M. Enabling big data education: teaching experience of big data education in Chinese colleges and universities[M]. Beijing: Publishing House of Electronics Industry, 2018.
- [2] 姚力,朱龙飞,崔晨. 大数据时代数据科学课程建设与人才培养的探索[J]. 计算机时代, 2018(11): 87-90, 93.  
YAO L, ZHU L F, CUI C. Exploration of data science course construction and personnel training in big data era[J]. Computer Era, 2018(11): 87-90, 93.
- [3] 耿秀荣. 数据科学与大数据技术专业建设的指导思想和基本措施[J]. 牡丹江教育学院学报, 2018(10): 16-17.  
GENG X R. Guiding ideas and basic measures for the construction of data science and big data technology specialty[J]. Journal of Mudanjiang College of Education, 2018(10): 16-17.
- [4] 吕小光,姜乐,成青松. 数据科学与大数据技术专业人才培养模式探析[J]. 淮海工学院学报(人文社会科学版), 2018, 16(9): 132-135.  
LYU X G, JIANG L, CHENG Q S. Research on the training model of data science and big data technology professionals[J]. Journal of Huaihai Institute of Technology, 2018, 16(9): 132-135.
- [5] 张志伟,房爱东,崔琳,等. 新工科背景下大数据专业建设探究[J]. 攀枝花学院学报, 2018, 35(5): 107-111.  
ZHANG Z W, FANG A D, CUI L, et al. Research on the construction of big data specialty under the background of new subject[J]. Journal of Pzhzhuhua University, 2018, 35(5): 107-111.
- [6] 张洁. 浅谈大数据导论的教学[J]. 计算机产品与流通, 2017(10): 234.  
ZHANG J. Discussion about introduction to big data[J]. Computer Products and Circulation, 2017(10): 234.
- [7] 张祖平. 数据科学与大数据技术导论[M]. 长沙: 中南大学出版社, 2018.  
ZHANG Z P. Introduction to data science and big data technology[M]. Changsha: Central South University Press, 2018.
- [8] 覃雄派,陈跃国,杜小勇. 数据科学概论[M]. 北京: 中国人民大学出版社, 2017.  
QIN X P, CHEN Y G, DU X Y. Introduction to data science[M]. Beijing: Renmin University Press, 2017.
- [9] 杨旭,汤海京,丁刚毅. 数据科学导论(第二版)[M]. 北京: 北京理工大学出版社, 2017.  
YANG X, TANG H J, DING G Y. Introduction to data science(V2)[M]. Beijing: Beijing University of Technology Press, 2017.
- [10] 刘鹏,张燕. 大数据导论[M]. 北京: 清华大学出版社, 2018.  
LIU P, ZHANG Y. Introduction of big data[M]. Beijing: Tsinghua University Press, 2018.
- [11] 周苏,王文. 大数据导论[M]. 北京: 清华大学出版社, 2016.  
ZHOU S, WANG W. Introduction of big data[M]. Beijing: Tsinghua University Press, 2016.

- [12] CORBEIL M E, CORBEIL J R, KHAN B H. A framework for identifying and analyzing major issues in implementing big data and data analytics in e-learning: introduction to special issue on big data and data analytics[J]. Educational Technology, 2017, 57(1): 3-9.

#### 作者简介



**张祖平** (1966-), 男, 博士, 中南大学教授、博士生导师、计算机科学与技术系主任, 大数据技术及应用团队负责人, 中南大学“531人才计划”第二层次人才, 中国计算机学会高级会员。加拿大西安大略大学(UWO)国家公派访问学者。2014年开始负责申报大数据专业的材料整理, 并在前期负责建设并管理专业。作为主要参与者申报《医疗大数据湖南省协同创新中心》与《医学大数据国家工程实验室》。长期从事大数据与知识工程、信息度量与信息融合、参数计算与生物计算等方向的研究。连续8年承担国家自然科学基金委员会基金项目, 项目推广应用到10多个省部级单位。发表学术论文100多篇, 作为主编出版了《信息学科导论》《数据库原理与应用》《计算机网络技术教程》等教材。主持国家自然科学基金项目面上项目2项、国家自然科学基金项目委主任基金项目2项, 企事业单位大型软件工程或信息化项目10多项, 多项大型软件系统通过国家或省级鉴定, 获省部级科技进步奖一等奖2项、二等奖1项、三等奖2项。目前正在承担长沙市大联合交管中心系统(交通大数据)建设, 负责《数据科学与大数据技术导论》教材、《面向大数据的形式概念分析及信息融合》专著

收稿日期: 2018-09-26