

目 录

第1章 专业概论	1
1.1 专业发展历史	1
1.1.1 专业产生背景	1
1.1.2 专业创办与申报	4
1.2 专业特点要求	4
1.2.1 专业培养定位	4
1.2.2 交叉型与复合型	4
1.3 专业课程模块	5
1.3.1 通识教育课	5
1.3.2 公共基础课	5
1.3.3 学科基础课	5
1.3.4 专业核心课	6
1.3.5 专业课	6
1.3.6 集中实践环节	6
1.3.7 专业完整课程体系供参考选择	7
1.4 专业技能体系	11
1.4.1 大数据环境	11
1.4.2 数据获取	11
1.4.3 数据处理与编程	11
1.4.4 数据挖掘与统计	15
1.4.5 数据预测	18
1.4.6 数据可视化	20
1.5 紧密相关的专业	23
1.5.1 计算机科学与技术	24
1.5.2 统计学	24
1.6 专业出路与就业	24
1.7 本章小结	25
思考题	25
本章相关的实验	26
第2章 数据科学与大数据基本概念	27
2.1. 数据相关的概念	27
2.1.1 基本概念	27

2.1.2 数据的分类	28
2.1.3 数据的属性	29
2.1.4 数据集	30
2.1.5 数据特征的统计描述	30
2.2 数据科学	37
2.2.1 数据科学定义	38
2.2.2 发展历史	40
2.2.3 研究内容	40
2.2.4 知识体系	40
2.2.5 与其他学科的关系	55
2.2.6 体系框架	55
2.3 基于数据科学的数据分析与挖掘	56
2.3.1 数据分析应用面临的挑战与发展	56
2.3.2 用好数据科学	57
2.3.3 数据科学平台工具	59
2.4 数据库	59
2.4.1 数据库概述	59
2.4.2 基本概念	60
2.4.3 数据库的分类	65
2.4.4 关系数据库系统操作语言	66
2.5 大数据	71
2.5.1 大数据定义及特征	71
2.5.2 大数据范式	71
2.6 本章小结	73
思考题	73
本章相关的实验	74
第3章 大数据核心技术	75
3.1 数据采集	75
3.1.1 软件接口方式	75
3.1.2 开放数据库方式	76
3.1.3 基于底层数据交换的数据直接采集方式	77
3.1.4 数据爬取	77
3.2 数据存储与管理	81
3.2.1 大数据存储与管理的主要模式	81

3.2.2 大数据存储典型三种技术路线	82
3.3 数据预处理	84
3.3.1 数据预处理的主要步骤	84
3.3.2 数据核查的主要方法	84
3.3.3 数据提质	85
3.3.4 数据集成	87
3.3.5 数据规约	87
3.3.6 数据变换	88
3.3.7 数据离散化	88
3.4 数据清洗	89
3.4.1 基本概念	90
3.4.2 数据清洗原理	90
3.4.3 需要清洗的主要数据类型	91
3.4.4 数据清洗方法	92
3.5 数据挖掘	92
3.5.1 起源	92
3.5.2 发展阶段	93
3.5.3 主要方法	94
3.5.4 行业应用	96
3.5.5 数据挖掘经典算法	96
3.5.6 关联规则挖掘	101
3.5.7 数据挖掘相关技术	101
3.6 数据可视化	101
3.6.1 概述	101
3.6.2 概念	102
3.6.3 主要应用	102
3.6.4 基本思想	102
3.6.5 基本手段	102
3.6.6 适用范围	103
3.6.7 发展阶段	104
3.6.8 大数据可视化	106
3.7 本章小结	106
思考题	107
本章相关的实验	107

第 4 章 大数据环境与技术	108
4.1 典型大数据环境及工具	108
4.1.1 Hadoop 综述	108
4.1.2 Hadoop 特点	109
4.1.3 Hadoop 核心架构	110
4.1.3.1 HDFS	148
4.1.4 Hadoop 的发展及社区服务	112
4.1.5 Hadoop 应用实例	114
4.1.6 Hadoop 安装	114
4.1.7 Hadoop 配置及启动服务	115
4.1.8 Hadoop 文件操作	118
4.2 典型大数据实用技术	120
4.2.1 存储 HDFS 及相关技术	120
4.2.2 计算 Yarn 及相关技术	125
4.2.3 计算 Spark 及相关技术	134
4.3 本章小结	140
思考题	140
本章相关的实验	141
第 5 章 大数据应用系统	142
5.1 医疗大数据	142
5.1.1 医疗大数据背景	144
5.1.2 医疗大数据应用技术研究中心	146
5.1.3 医疗大数据应用关键技术	167
5.1.4 引领未来的关键共性技术	169
5.1.5 医疗大数据软硬件环境	172
5.2 交通大数据	172
5.2.1 交通大数据背景	172
5.2.2 交通大数据应用中面临的问题	173
5.2.3 交通大数据数据特点及数据来源	174
5.2.4 交通大数据融合技术	175
5.2.5 交通大数据的全流程分层次特点与技术	182
5.2.6 交通大数据安全技术	183
5.2.7 交通大数据的数据发现	185
5.2.8 交通管理数据库设计技术	187

5.2.9 交通大数据应用	189
5.2.10 交通大数据软硬件环境	193
5.2.11 交通大数据分析与展示技术	196
5.3 本章小结	199
思考题	199
本章相关的实验	200
附录：数据科学与大数据技术培养方案	201
参考文献	217