



第4届全国高校大数据与人工智能教学研讨会

2021.05.14-05.15 中国·厦门

主办单位：教育部高等学校计算机类专业教学指导委员会

承办单位： 厦门大学
 山东大学
 华东师范大学
 清华大学
 北京大学
 集美大学
 湖南师范大学
 人民邮电出版社

协办单位： 头歌 EduCoder
 TRANSWARP 星环科技
 山东省计算中心
 荔峰科技 LIFENG
 探知图灵 TOUZHIDATOU
 美林数据 MERITDATA

数据科学与工程专业建设 成果与经验



周烜

xzhou@dase.ecnu.edu.cn

华东师范大学



华师大的数据专业建设过程

3

□ 2012.6 云计算与大数据研究中心

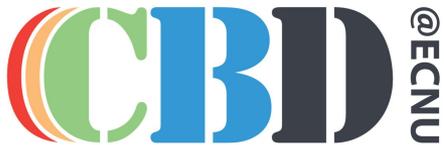
- ⊙ 研究生培养

□ 2013.9 数据科学与工程研究院

- ⊙ 全国高校第一个大数据研究院
- ⊙ 学科交叉，应用驱动创新
- ⊙ 完善“数据科学与工程”二级学科

□ 2016.9 数据科学与工程学院

- ⊙ 2017：招收“数据科学与大数据技术”专业本科生
- ⊙ 2017：启动**CST**计划（Curriculum-Syllabus-Textbook）
- ⊙ 2019：获评国家级一流本科
- ⊙ 2020：第一届本科生毕业



华东师范大学云计算与大数据研究中心
Center for Cloud Computing and Big Data @ ECNU



数据科学与工程专业协作组

4

- 华东师范大学
- 中国人民大学
- 东北大学
- 哈尔滨工业大学
- 西北工业大学
- 中山大学
- 中国矿业大学
- 河南大学
- 湖北大学
- 石河子大学
- 中原工学院
- 桂林电子科技大学
- 贵州大学
- 云南大学
- 四川大学
- 优达学城
- 高等教育出版社

协作组的工作历程

5

2018年12月25日
第一次研讨会 + 成立
仪式
华东师范大学



2019年5月11-12日
第二次研讨会
河南大学



2019年7月23日
第三次研讨会
华东师范大学

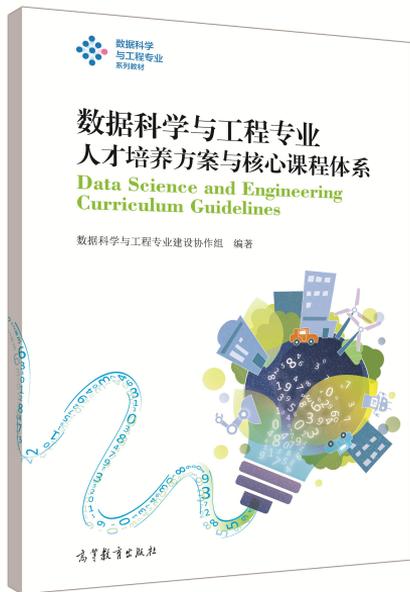


2019年11月4日
第四次研讨会
中山大学



初步成果

6

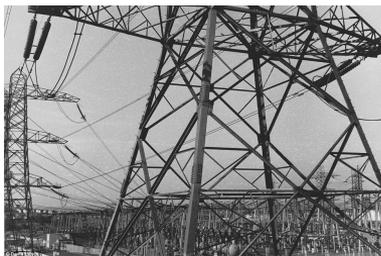


- 数据科学与工程专业的培养目标
 - ◎ “数据科学与大数据技术”专业
- 数据科学与工程专业的课程体系
- 数据科学与工程专业核心课程大纲
- 系列教材（进行中）

数据是新的电能

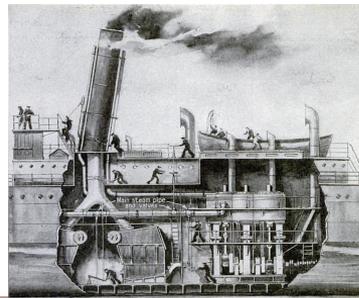
7

1880s-1990s
electric power
电能



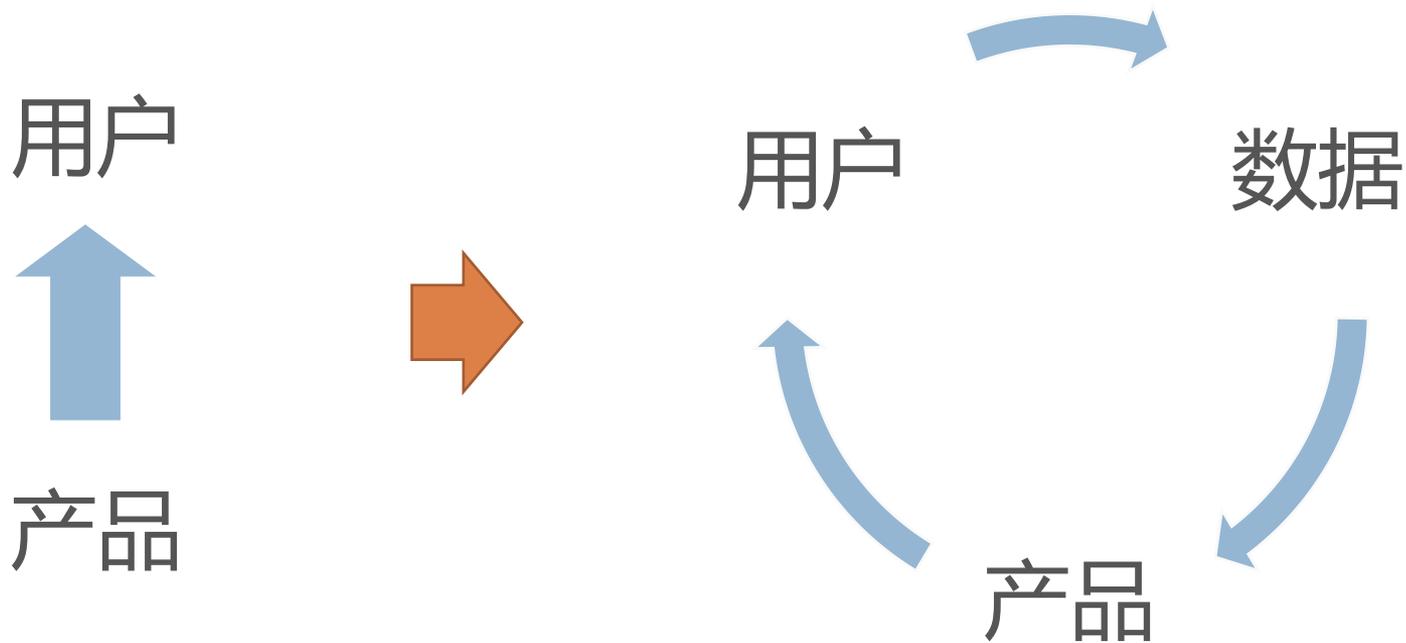
2000s-未来
data power
数据能（新生产要素）

1760s-1870s
steam power
蒸汽能



数据赋能的生产方式

8





数据

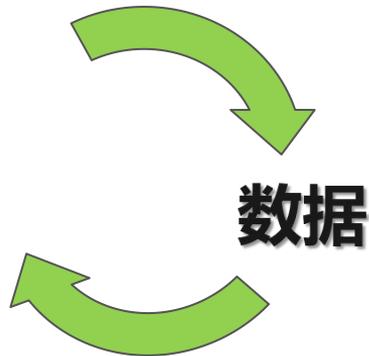




数据专业的人才培养目的

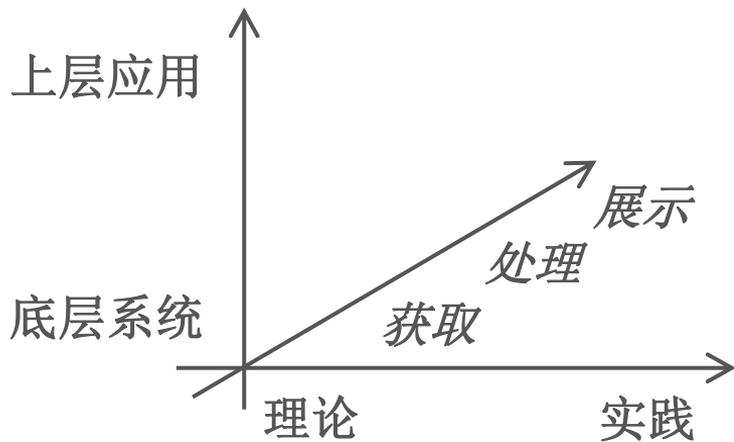
12

- 培养学生“数据赋能”的能力
 - ⊙ 数据思维
 - ⊙ 数据分析与处理技能
 - ⊙ 用数据驱动业务的能力



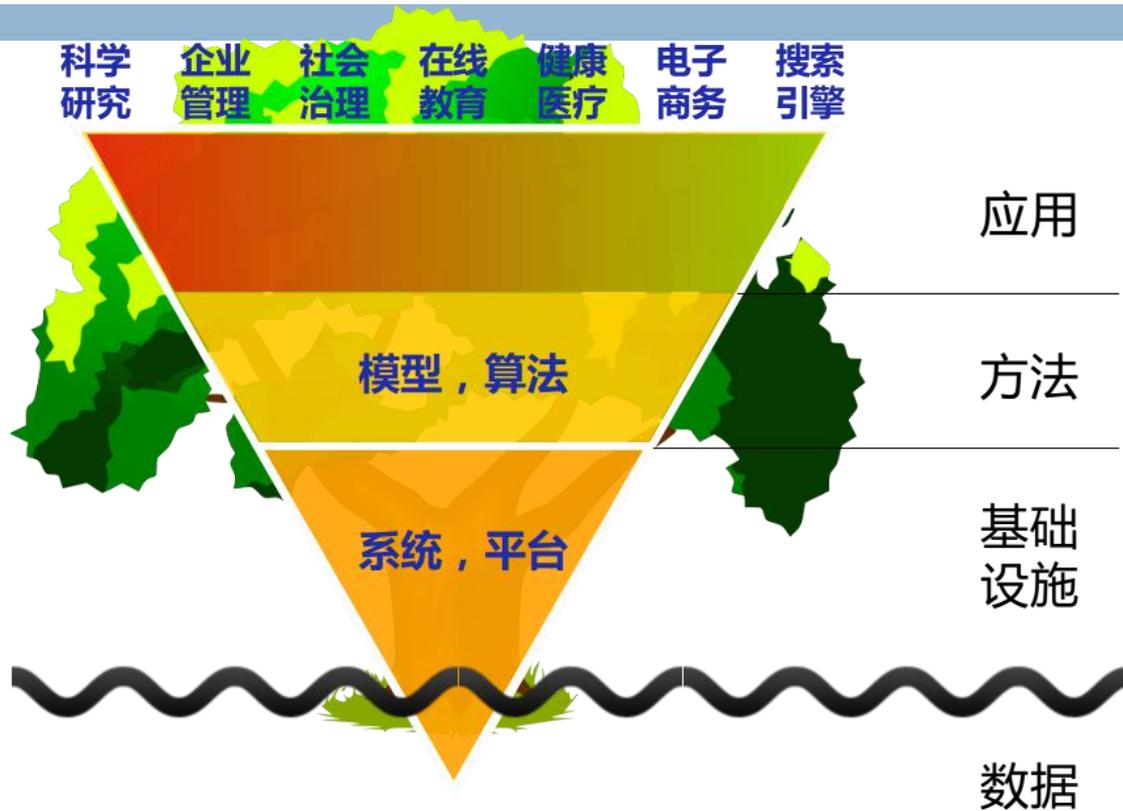
数据专业的知识体系

13



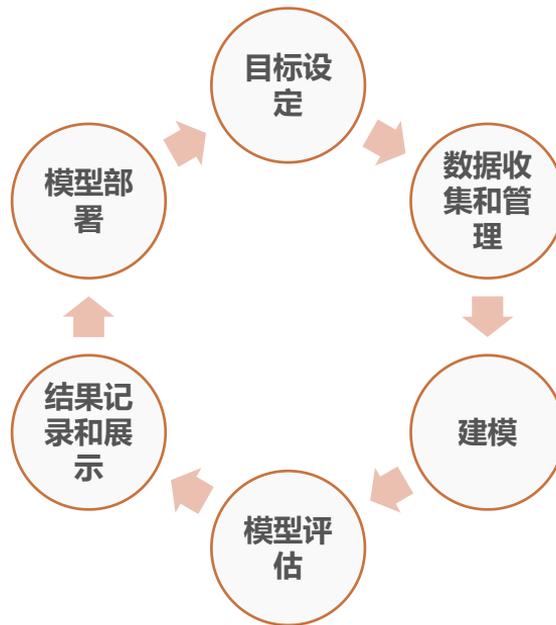
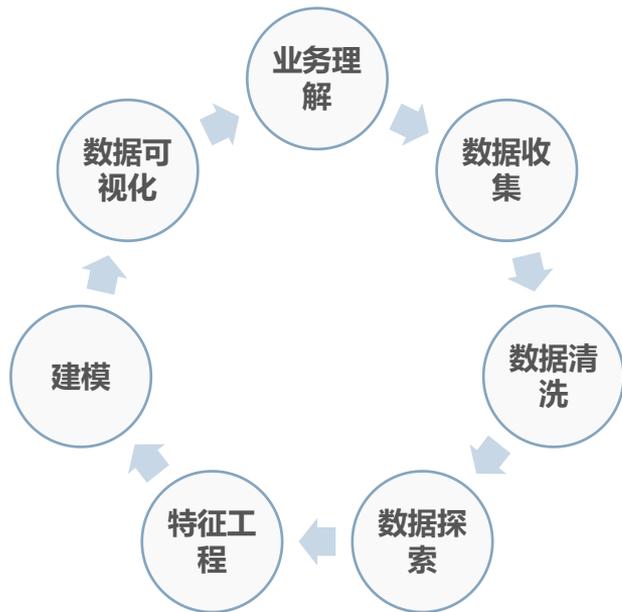
数据系统的体系

14



数据科学的生命周期 - 过程

15



理论到实践

16

□ 理论知识

- ◎ 数学基础
- ◎ 统计理论
- ◎ 计算机理论
- ◎ 算法基础
- ◎ AI理论

□ 工程实践能力

- ◎ 程序设计
- ◎ 软件开发
- ◎ 架构设计
- ◎ 系统工具的使用
- ◎ 系统工具的构建

□ 业务实践能力

- ◎ 对应用的理解
- ◎ 数据价值评估
- ◎ 数据驱动业务
- ◎ 设计思维
- ◎ 数据伦理

课程体系

17

□ 基础课

- ◉ 高等数学、线性代数、概率论与数理统计、程序设计、数据结构、算法、计算机系统、操作系统、计算机网络

□ 核心课

- ◉ **数据科学与工程导论、数据科学与工程**的数学基础、**数据科学与工程**的算法基础、**云计算系统、当代数据**管理系统、**分布式计算系统、统计方法与机器学习、当代人工智能**

□ 方向课（选修）

- ◉ 数据安全与隐私、数据挖掘、深度学习、自然语言处理、数据治理、机器视觉、数据中台、金融科技、智慧营销、智慧城市、商业分析、区块链

□ 实践课

- ◉ Web应用开发、开源实践与软件工程、数据可视化、数据伦理、设计思维、毕业实习（项目实践）

基础课

高等数学
线性代数
~~离散数学~~
概率论与数理统计

程序设计
数据结构
算法设计与分析

计算机系统
操作系统
计算机网络

核心课

数据科学与工程导论
数据科学与工程的数学基础
数据科学与工程的算法基础

云计算系统
当代数据管理系统
分布式计算系统

统计方法与机器学习
当代人工智能

方向课

数据挖掘
深度学习
自然语言处理
计算机视觉
多媒体技术
信息检索
数据治理
高级数据库系统
社会计算
商业分析
区块链
金融大数据
物流大数据
教育大数据
...

工程实践课

Web编程、数据可视化、软件工程、开源软件实践、设计思维、数据伦理、毕业设计（毕业实习）

课程安排

	第1学期	第2学期	第3学期	第4学期	第5学期	第6学期	第7学期	第8学期
基础课	高等数学(上) 线性代数 程序设计 数据结构	高等数学(下) 概率论与数理统计 计算机系统	算法设计与分析 操作系统	计算机网络				
核心课			数据科学与工程导论 数据科学与工程 的数学基础	数据科学与工程 的算法基础 云计算系统	当代数据管理 系统 统计方法与机 器学习 当代人工智能	分布式计算 系统		
实践课		Web编程		开源软件与软 件工程	数据可视化	数据伦理	设计思维	毕业设计
方向课 (选修)						5选2	5选2	5选2

数据科学与工程导论

20

- 时间：二年级上
- 学分：2-3
- 先修课：高等数学、线性代数、程序设计
- **目的：训练学生的数据思维，让学生了解专业的目的，体会数据科学家和数据工程师的工作**

数据科学与工程导论

21

知识点	学时
数据科学与工程的基本概念	3
数据思维与问题求解	3
数据的模型与结构	3
数据的计算与程序表达	3
计算基础设施	3
数据的全生命周期管理	3
数据库系统	3

知识点	学时
大数据系统	3
数据科学过程	3
统计分析原理	3
机器学习	6
数据挖掘	6
数据综合应用	3
数据道德与职业	3

数据科学与工程学院

《基于交通信息参数的探索和分析》

数据科学与工程学院

■ 学校名称: 华东师范大学
■ 讲师: 鞠磊
■ 邮箱: julei@sjtu.edu.cn

基于NBA球员数据的分析与建模

上海文化空间分析

×
REPORTER: 张双宇
TIME: 2019.1.10

基于房屋价格数据的分析与建模

甄若男

基于小米电视评论文本数据的情感分析
10172100347
史琳琳
2019年1月

爬取《我不是药神》影评进行可视化展示



金庸的武侠世界

帕金森症的预测

演示者 王康超



用k-means改进knn在cifar-10上的预测时间

SymbolNet : First Step of Handwriting to Latex

Tao Cheng

East China Normal University
taocheng02@gmail.com

January 10, 2019

图像文本识别模型探索

王子刚 10175501108

基于线性回归和岭回归的台风路径预测

魏加波

基于Python工作的数据分析与可视化

基于北美进口影片在中国票房表现数据的分析与建模



基于上海市二手房数据的分析

数据科学与大数据技术
徐宇博

© 2019 2017 版本科学 ID: 10175501117

《数据科学与工程导论》

23

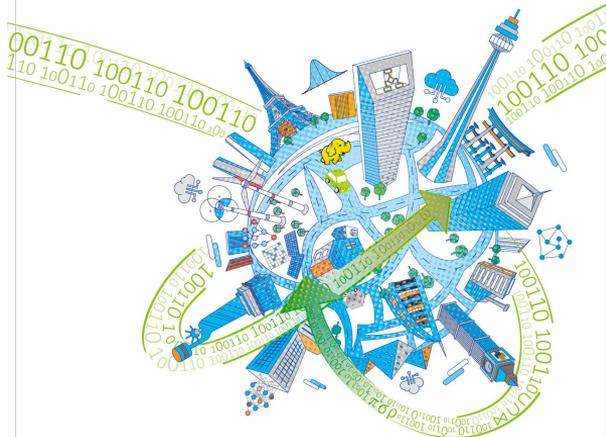


数据科学
与工程专业
系列教材

数据科学与工程导论

Introduction to Data Science and Engineering

王伟 刘垚 编著



高等教育出版社
HIGHER EDUCATION PRESS

教材已出版

作者：王伟、刘垚

华东师范大学出版社

数据科学与工程数学基础

24

- 时间：二年级上
- 学分：3-4
- 先修课：高等数学、线性代数、概率论、数理统计、离散数学
- **目的：补充专业所需的重要数学知识，包括矩阵分析、信息论、优化理论**

数据科学与工程数学基础

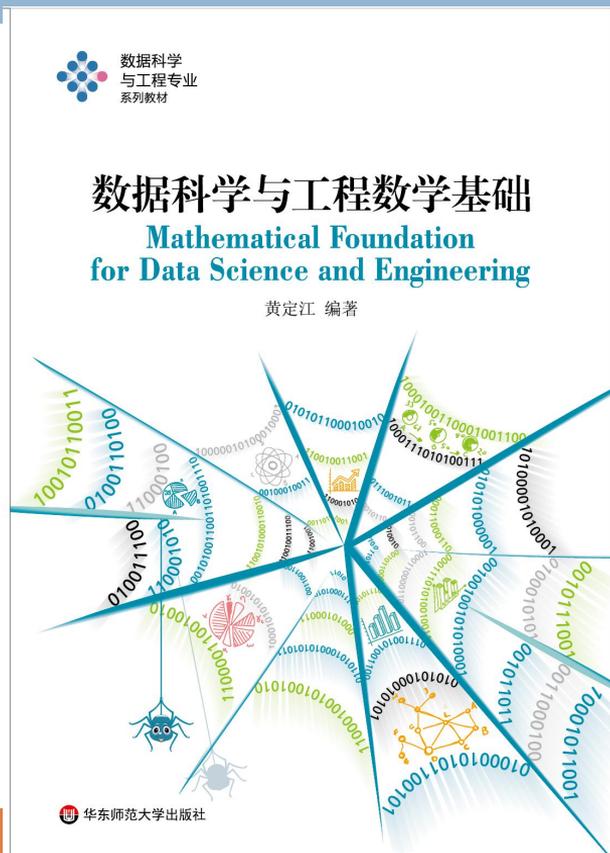
25

知识点	学时
课程简介、数据科学与工程数学基础介绍	2
向量与矩阵基础	9
矩阵分解	7
矩阵计算问题	8
向量与矩阵微分	5

知识点	学时
概率基础	6
信息论基础	3
概率模型与参数估计	6
凸优化基础	7
最优性和对偶性	6
优化算法	8

《数据科学与工程数学基础》

26



教材正在出版

作者：黄定江

华东师范大学出版社

数据科学与工程算法基础

27

- 时间：二年级下
- 学分：2-3
- 先修课：高等数学、线性代数、概率论、数理统计、离散数学、数据科学与工程数学基础、数据结构与算法
- **目的：培养数据处理所需的算法设计能力**

数据科学与工程算法基础

28

知识点	学时
绪论：数据相关的算法设计	3
抽样算法	4
随机算法基础	4
哈希算法	4
数据流算法	4
整数规划	3

知识点	学时
子模及其应用	4
特征值计算	3
SVD和PCA	4
矩阵分解算法	6
随机游走算法	4
社团发现算法	4
EM算法	4

数据科学与工程算法基础

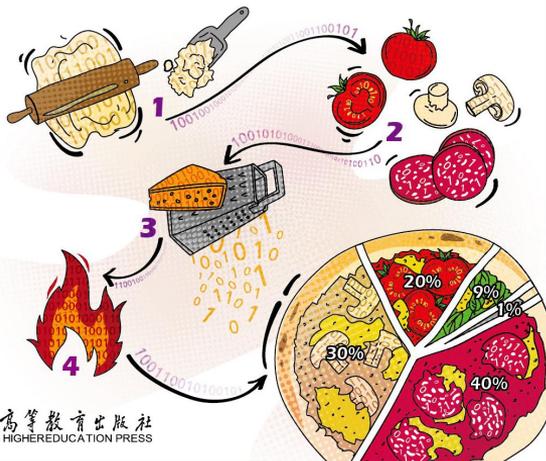
29



数据科学
与工程专业
系列教材

数据科学与工程算法基础 Algorithm Foundations for Data Science and Engineering

高明 胡卉芪 编著



教材正在出版

作者：高明、胡卉芪

高等教育出版社

云计算系统

30

- 时间：二年级下
- 学分：2
- 先修课：程序设计、计算机系统
- **目的：培养使用云计算平台构建数据应用的能力**

云计算系统

31

知识点	学时
云计算基础	4
云计算核心技术 (虚拟化、分布式)	4
云服务的整体开发流程 (云原生开发流程与框架)	4
云服务的基础设施构建 (微服务、容器)	6

知识点	学时
云服务的编排与管理 (Kubernetes)	6
云服务的运维 (DevOps、Prometheus)	6
云计算进阶技术 (Serverless、FaaS)	4

当代数据管理系统

32

- 时间：三年级上
- 学分：3-4
- 先修课：离散数学、程序设计、操作系统、Web编程
- **目的：培养使用各类数据管理系统的能力**

当代数据管理系统

33

知识点	学时
数据管理系统的概念	3
数据模型	4
文档数据库系统	7
文档数据库设计	5
关系数据库系统	10
关系数据库设计	9

知识点	学时
事务处理	9
数据库的维护与调优	5
数据管理系统的扩展	8
数据分析系统	9
其他数据管理系统	5

分布式计算系统

34

- 时间：三年级下
- 学分：2-3
- 先修课：程序设计、计算机系统、操作系统、云计算原理和计算、当代数据管理系统
- **目的：培养使用各类分布式数据处理平台的能力**

分布式计算系统

35

知识点	学时
分布式计算系统简介	4
分布式文件系统HDFS	4
批处理系统MapReduce	5
批处理系统Spark	6
资源管理系统Yarn	3
协调服务系统ZooKeeper	6

知识点	学时
流计算系统Storm	4
流计算系统Spark Streaming	4
批流融合概述	6
批流融合系统Flink	8
图处理系统Giraph	4
机器学习系统SystemML	3

统计方法与机器学习

36

- 时间：三年级上
- 学分：3-4
- 先修课：高等数学、线性代数、概率论、数理统计、离散数学、数据科学与工程中的数学基础
- **目的：培养学生数据建模的能力**

统计方法与机器学习

37

知识点	学时
方差分析	7
回归分析	8
变量选择	10
聚类方法	5
判别与分类方法	8

知识点	学时
支持向量机	10
人工神经网络	10
决策树	2
集成学习	4

当代人工智能

38

- 时间：三年级上
- 学分：3-4
- 先修课：离散数学、概率论、数理统计、算法设计与分析、数据科学与工程_的数学基础
- **目的：培养学生使用数据构建人工智能应用的能力**

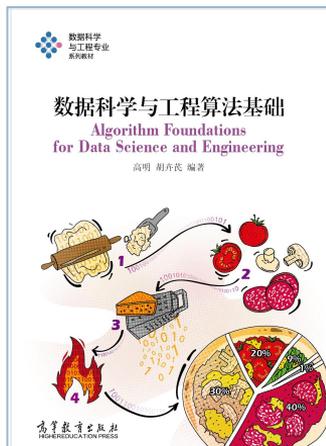
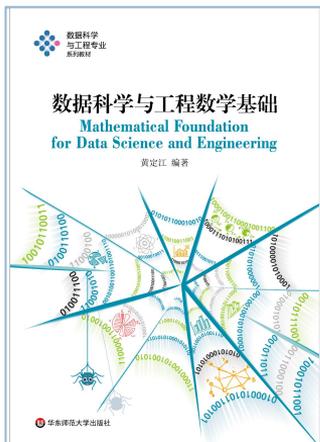
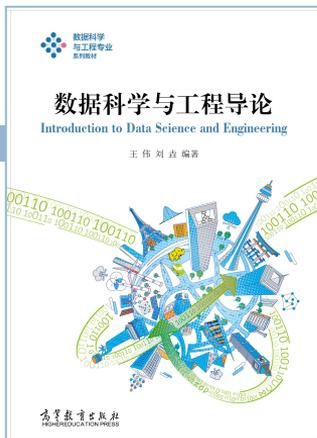
当代人工智能

39

知识点	学时
人工智能发展历程	3
图像目标检测中的人工智能技术	15
机器翻译中的人工智能技术	12
博弈游戏中的人工智能技术	16
基于知识的问答系统中的人工智能技术	15
人工智能综合应用	3

教材规划

40



敬请期待

- 《当代数据管理系统》2021
- 《分布式计算系统》2021
- 《云计算系统》2021
- 《统计方法与机器学习》2021
- 《当代人工智能》2022

水杉平台

41



水杉学堂

复杂报表
自然语言处理
分析计算
多维分析

查看详情



水杉工坊

复杂报表
自然语言处理
分析计算
多维分析

查看详情



水杉校场

复杂报表
自然语言处理
分析计算
多维分析

查看详情



水杉码园

复杂报表
自然语言处理
分析计算
多维分析

查看详情

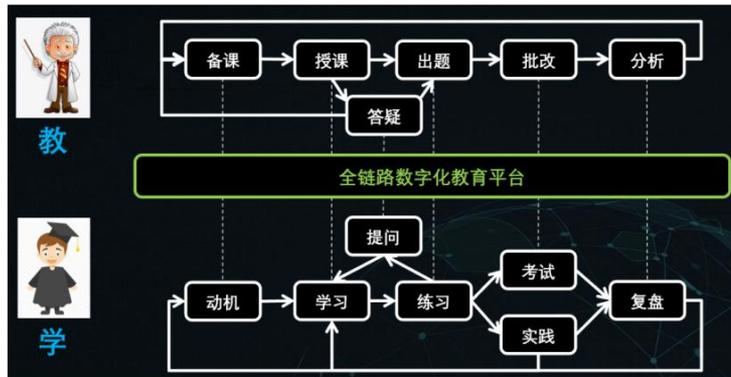
以编程思维的养成为核心：

“水杉学堂”：开放式全民计算机科学教育课堂；

“水杉工坊”：交互式在线实训服务的极速体验；

“水杉校场”：个性化在线编程学习的自动评测；

“水杉码园”：社交化软件项目创作的协作分享。



Thanks!

42

