
《大数据及机器学习算法》

——段方

某世界 100 强企业大数据总设计师
教授 北京大学博士后

1 概述

1.1 大数据的概念和特点

1.1.1 大数据的基础

1.1.2 大数据如何“与时俱进”？

1.1.3 大数据发展趋势

[1.1.3.1 人工智能](#)

[1.1.3.2 物联网结合](#)

[1.1.3.3 各个行业的深入](#)

1.2 大数据的技术基础

1.2.1 从数据仓库开始

1.2.2 HADOOP 生态圈

1.2.3 与云计算的关系

1.2.4 数据运维能力提升

1.3 大数据的应用举例

1.3.1 大数据提升客户分析能力

- 1.3.2 大数据提升产品分析能力
- 1.3.3 大数据提升管理水平
- 1.3.4 大数据提升各行业“智慧 ”
- 1.4 大数据下的人工智能 (AI)
- 1.4.1 什么是人工智能
- 1.4.2 人工智能改变哪些行业？
- 1.4.3 大数据下的人工智能有何不同？
- 1.4.4 人工智能的“颠覆 ”
- 1.5 大数据如何精细化管理
- 1.5.1 量化管理的引出
- 1.5.2 大数据如何提升“量化 ”的维度和深度
- 1.5.3 从艺术到技术
- 1.5.4 自动驾驶到自动管理？
- 1.6 金融企业的大数据“商机 ”
- 1.6.1 从网络运营到数据运营
- 1.6.2 提炼“ 内功 ”
- 1.6.3 提升外部管理能力
- 1.6.4 扩展增值产品运营市场

2 机器学习基础

2.1 多维分析方法

2.1.1 OLAP 分析

2.1.2 上钻和下钻

2.1.3 用 OLAP 分析问题

2.2 分析算法

2.2.1 回归算法

[2.2.1.1 线性回归](#)

[2.2.1.2 逻辑回归](#)

2.2.2 决策树算法

[2.2.2.1 C4.5 算法](#)

[2.2.2.2 CART 算法](#)

2.2.3 贝叶斯算法

[2.2.3.1 朴素贝叶斯算法](#)

[2.2.3.2 BBN \(Bayesian Belief Network\) 算法](#)

2.2.4 基于核的算法

[2.2.4.1 支持向量机 SVM 算法](#)

[2.2.4.2 线性判别分析\(Linear Discriminate Analysis , LDA\)](#)

2.2.5 聚类算法

[2.2.5.1 K-MEANS 算法](#)

[2.2.5.2 期望最大化算法\(Expectation Maximization , ME\)](#)

2.2.6 关联规则算法

[2.2.6.1 Apriori 算法](#)

2.2.7 降低维度算法

[2.2.7.1](#) 主成份分析(Principle Component Analysis , PCA)算法

[2.2.7.2](#) 偏最小二乘回归(Partial Least Square Regression , PLS)算法

2.2.8 集成算法

[2.2.8.1](#) 随机森林算法

[2.2.8.2](#) 梯度推进机

2.3 机器学习方法在金融行业应用举例

3 大数据技术基础

3.1 从数据仓库开始

3.1.1 数据仓库的“集中”

3.1.2 数据仓库的模型标准化

3.1.3 大数据的演进

3.2 HADOOP 生态圈

3.2.1 开源社区概述

3.2.2 开源改变了什么？

3.2.3 HADOOP 生态圈内容

3.2.4 HADOOP 的技术原则

3.2.5 HADOOP 的运维

3.3 HADOOP 基础

3.3.1 HDFS 的原理

3.3.2 MAP/REDUCE 原理

3.3.3 YARN 原理

3.4 HIVE/ HBASE 技术

3.4.1 HIVE 的原理

3.4.2 HBASE 的原理

3.4.3 两者的关系

3.5 SPARK 技术

3.5.1 基本原理

3.5.2 应用案例

3.6 KAFKA/ FLUME

3.6.1 基本原理

3.6.2 应用案例

3.7 各个技术的特点对比

4 大数据下的企业运营

4.1 从泰勒经济学说起

4.1.1 简述泰勒经济学

4.1.2 数据与企业管理的天然联系

4.1.3 让大数据渗透到企业的每个毛孔

4.2 客户分析

4.2.1 客户的数据维度

4.2.2 客户视图分析

4.2.3 客户离网分析等

4.3 产品分析

4.3.1 产品数据维度

4.3.2 产品视图分析

4.3.3 产品的潜在客户分析

4.3.4 如何优化产品？

4.4 精准营销

4.4.1 营销的本质——配对

4.4.2 精准营销——数据的魅力

4.4.3 营销时机、渠道等选择

4.5 员工的管理

4.5.1 员工的量化分析

4.5.2 大数据下的“办公室分析”

4.5.3 员工绩效考评

4.5.4 如何避免“苍蝇犯大案”

4.6 外部生态圈的管理

4.6.1 生态圈的概念

4.6.2 金融企业的生态圈

4.6.3 上下游企业的管理

4.6.4 如何识别渠道欺诈？

4.7 【例】某大数据系统应用案例

5 *大数据的互联网思维

5.1 互联网思维概述

5.1.1 什么是互联网思维？

5.1.2 互联网思维在互联网企业中的落地

5.1.3 互联网思维的案例

5.2 大数据的客户体验

5.2.1 什么是客户体验？

5.2.2 大数据的客户有哪些？

5.2.3 如何提升客户大数据使用体验？

5.2.4 【例】大数据的客户提升案例

5.3 大数据的产品设计

5.3.1 大数据产品应该长什么样？

5.3.2 产品如何简约？

5.3.3 提供“千人千面”的产品

5.3.4 大数据的“爆品”在哪里？

5.3.5 【例】大数据产品示例

5.4 大数据的极致思维

5.4.1 如何极致？

5.4.2 数据的极致

5.4.3 报告的极致

5.4.4 讲故事的极致

5.4.5 服务的极致

5.5 大数据的平台思维

5.5.1 什么是平台思维？

5.5.2 大数据的平台化如何实现？

5.5.3 从 APP STORE 到 BI STORE

5.6 其它

5.6.1 互联网思维如何用于企业运营？

5.6.2 互联网思维堪比“文艺复兴”？

5.7 【例】互联网思维在某企业大数据系统中的落地实践

6 大数据的收集及整理

6.1 大数据的收集

6.1.1 内部数据

[6.1.1.1 哪些 IT 系统？](#)

[6.1.1.2 财务系统](#)

[6.1.1.3 招聘系统等](#)

6.1.2 外部数据

[6.1.2.1 爬虫获取数据](#)

[6.1.2.2 交换数据](#)

[6.1.2.3 购买数据等](#)

6.2 大数据的整理

- 6.2.1 如何表述数据间的关系？
- 6.2.2 数据建模的概念
- 6.2.3 数据建模的方法
- 6.3 大数据的数据治理
 - 6.3.1 数据质量问题
 - 6.3.2 数据质量如何管控
 - 6.3.3 数据字典的定义
- 6.4 大数据的数据资产
 - 6.4.1 什么是数据资产？
 - 6.4.2 数据资产如何梳理？
- 6.5 【案例】某数据收集整理实际案例

7 金融行业大数据案例

- 7.1 某金融企业大数据建设案例分享
- 7.2 某金融企业大数据应用案例分享
- 7.3 某金融企业大数据管理案例分享
- 7.4 某金融企业大数据变现案例分享
- 7.5 其它行业大数据案例分享

8 总结
