
《数据分析可视化与 AI 算法实践》

-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 背景

1.1 从 AlphaGo 说起

1.1.1 AlphaGo 的效果

1.1.2 AlphaGo 的原理

1.2 机器学习基础

1.2.1 机器学习的历史

1.2.2 机器学习概念

1.2.3 机器学习与数据挖掘

1.3 深度学习基础

1.3.1 深度学习的概念和特点

1.3.2 深度学习的意义

1.3.3 深度学习的应用领域

1.4 应用与技术的平衡

1.4.1 数据驱动还是应用驱动？

1.4.2 应用价值的显现化

1.4.3 对内服务和对外服务

1.5 GPU 的引出

1.5.1 计算的硬件基础

1.5.2 与 CPU 的对比

1.5.3 与 FPGA 的对比

1.5.4 GPU 的示例

1.6 建设应用的生态圈——GitHub

1.6.1 为什么要建生态圈

1.6.2 生态圈的构建方法

1.6.3 GITHUB 的意义

1.7 【案例】附件——大数据和 AI 对各个行业的影响

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 【案例】附件-机器学习方法在金融行业应用举例

3 数据分析的可视化

3.1 可视化的概念和范畴

3.2 从统计分析开始的可视化

3.2.1 Excel 的可视化

3.3 传统数据分析工具的可视化

3.3.1 SPSS 分析的可视化

3.4 开源的可视化工具

3.4.1 百度 Echarts 的可视化

3.4.2 阿里 DataV 的可视化

3.5 图表的可视化美化

3.5.1 canva

3.6 【案例】附件——数据分析可视化案例

4 人工智能的应用内容

4.1 模式识别应用

4.1.1 人脸识别

4.1.2 文字识别

4.1.3 物体识别

4.2 博弈类应用

4.2.1 IBM 的“深蓝”

4.2.2 “紫光之星”

4.3 专家系统

4.3.1 医疗领域

4.3.2 探矿领域

4.4 机器人

4.4.1 KAKU 机器人

4.4.2 机器人足球

4.4.3 机器人舞蹈

4.5 机器视觉

4.5.1 图像识别

4.5.2 罪犯人脸特征分析

4.6 自然语言理解

4.6.1 谷歌翻译

4.6.2 语音识别

4.7 其它应用

4.7.1 自动程序设计

4.7.2 智能信息检索

4.7.3 声纹识别

4.7.4 智能仿真

4.8 【案例】AlphaGo 下围棋的原理介绍

5 人工智能的深度学习

5.1 概述

5.1.1 为什么是深度学习？

[5.1.1.1](#) 引出

[5.1.1.2](#) 与浅层学习的区别

[5.1.1.3](#) 原因

5.1.2 什么是“无监督”学习？

5.1.3 与神经网络的关系？

5.2 神经网络学习

5.2.1 概念

[5.2.1.1](#) 脑神经元分析

5.2.2 原理

[5.2.2.1](#) BP 网络

5.3 深度学习介绍

5.3.1 多层神经网络

5.3.2 深度学习的弱点

[5.3.2.1](#) 缺乏时间概念

[5.3.2.2](#) 视频与图片

5.4 深度学习原理

5.4.1 从单层神经网络到多层神经网络

5.4.2 深度学习的训练过程

5.4.3 深度学习的具体模型及方法

5.4.4 深度学习的性能比较

5.4.5 深度学习的应用

5.5 深度学习的意义

5.5.1 改变了传统人工智能的哪些思维定式？

5.5.2 深度学习的无监督学习

5.6 TensorFlow

5.6.1 技术框架

5.6.2 与 Python 的关系

5.6.3 TensorFlow 实现卷积神经网络

5.6.4 TensorFlow 实现循环神经网络

5.6.5 TensorFlow 实现多 GPU 并行及分布计算

5.7 【案例】某行业人工智能应用案例

6 人工智能的算法解析

6.1 人工智能的算法范畴

6.1.1 搜索算法

6.1.2 博弈算法

6.1.3 模糊算法

6.1.4 遗传算法

6.2 机器学习算法

6.2.1 C4.5 算法

6.2.2 K-means 算法

6.2.3 朴素贝叶斯算法

6.2.4 K 最近邻分类算法

6.2.5 ME 最大期望算法

6.2.6 PAGERANK 算法

6.2.7 AdaBoost 算法

6.2.8 APRIORI 算法

6.2.9 CART 分类与回归树

6.3 深度学习的算法

6.3.1 (多层) 感知机

6.3.2 深度神经网络 (DNN)

6.3.3 循环神经网络 (RNN)

6.3.4 卷积神经网络 (CNN)

6.3.5 长短期记忆网络 (LSTM)

6.4 数据分析框架——Spark MLlib

6.4.1 Spark MLlib 的基础框架与原理

6.4.2 Spark MLlib 基本算法

[6.4.2.1](#) Spark MLlib 支持的分类算法

[6.4.2.2](#) Spark MLlib 支持的聚类算法

[6.4.2.3](#) Spark MLlib 协同过滤

6.4.3 Spark MLlib 中的矩阵向量运算库 jblas

6.4.4 Spark MLib 梯度下降算法

6.4.5 【实际操作】Spark MLib 中 K-means 算法源码相关参数、分析与调试，熟悉 k_means 的步骤

[6.4.5.1 K-means 算法源码分析](#)

[6.4.5.2 K-means 算法实例](#)

6.4.5.2.1 数据准备及参数配置

6.4.5.2.2 使用 KMeans 建立模型、训练、预测结果及模型评估

6.5 【案例】算法实际应用案例

6.5.1 客户语音识别案例

6.5.2 人脸识别案例

7 人工智能的应用案例：自然语言处理

7.1 NLP 的技术原理

7.2 "大模型、多模态"的代表——GPT-3

7.3 聊天机器人的代表——ChatGPT

7.4 【案例】ChatGPT 举例

8 人工智能的关键点

8.1 数据重要还是算法重要 ?

8.2 如何收集非结构化数据 ?

8.3 应用的最后一公里问题

8.4 人工智能应用步骤规划

8.5 如何借助他人的“肩膀”？

9 总结
