

《基于大数据的机器学习和深度学习》

——原理与实践

目录

1 背景.....	5
1.1 从 AlphaGo 说起	5
1.1.1 AlphaGo 的效果.....	5
1.1.2 AlphaGo 的原理.....	5
1.2 机器学习基础.....	5
1.2.1 机器学习的历史.....	5
1.2.2 机器学习概念	5
1.2.3 机器学习与数据挖掘.....	5
1.3 深度学习基础.....	5
1.3.1 深度学习的概念和特点.....	5
1.3.2 深度学习的意义.....	5
1.3.3 深度学习的应用领域.....	5
1.4 应用与技术的平衡.....	5
1.4.1 数据驱动还是应用驱动？	5
1.4.2 应用价值的显现化	5
1.4.3 对内服务和对外服务.....	6
1.5 GPU 的引出.....	6
1.5.1 计算的硬件基础.....	6
1.5.2 与 CPU 的对比.....	6
1.5.3 与 FPGA 的对比.....	6
1.5.4 GPU 的示例	6
1.6 建设应用的生态圈.....	6
1.6.1 为什么要建生态圈	6
1.6.2 生态圈的构建方法	6
2 大数据应用开发方法.....	6
2.1 深度学习的方法论有何异同？	6
2.2 CRISP-DM (CRoss-Industry Standard Process for Data Mining) 方法论.....	6
2.3 业务理解 (Business Understanding)	6
2.4 数据理解 (Data Understanding)	6
2.5 数据准备 (Data Preparation)	6
2.6 建模 (Modeling)	7
2.7 评估 (Evaluation)	7
2.8 部署 (Deployment)	7
2.9 数据挖掘之外的方法论.....	7
3 机器学习基础.....	7
3.1 多维分析方法.....	7
3.1.1 OLAP 分析.....	7
3.1.2 上钻和下钻.....	7
3.1.3 用 OLAP 分析问题.....	7
3.2 分析算法	7
3.2.1 回归算法.....	7
3.2.2 决策树算法.....	7

3.2.3 贝叶斯算法.....	8
3.2.4 基于核的算法	8
3.2.5 聚类算法.....	8
3.2.6 关联规则算法	8
3.2.7 降低维度算法	8
3.2.8 集成算法.....	9
3.3 机器学习方法在电信行业应用举例.....	9
4 深度学习基础.....	9
4.1 神经网络算法的演进历史.....	9
4.1.1 传统的神经网络算法.....	9
4.1.2 神经网络算法的特点.....	9
4.1.3 神经网络算法的瓶颈点.....	9
4.2 神经网络的算法原理	9
4.2.1 神经元的原理	9
4.2.2 神经网络的原理.....	9
4.2.3 神经网络的 BP 算法	9
4.3 多层神经网络算法.....	9
4.3.1 单层神经网络	9
4.3.2 两层神经网络	10
4.3.3 多层神经网络——深度学习.....	10
4.4 深度学习算法.....	10
4.4.1 卷积神经网络 CNN.....	10
4.4.2 循环神经网络 RNN.....	10
4.4.3 深度神经网络 DNN	10
4.5 深度学习的应用案例	10
4.5.1 深度学习用于图像识别.....	10
4.5.2 深度学习用于语音识别.....	10
4.5.3 深度学习用于语言翻译.....	10
5 TensorFlow 等框架.....	10
5.1 人工智能的学习框架	10
5.1.1 开源的概念.....	10
5.1.2 学习框架的历程.....	10
5.1.3 caffe 等学习框架介绍	10
5.2 TensorFlow 框架内容.....	11
5.2.1 编程模型简介	11
5.2.2 自编码器介绍	11
5.2.3 多层感知机介绍.....	11
5.3 TensorFlow 支持的算法.....	11
5.3.1 支持卷积神经网络算法.....	11
5.3.2 支持循环神经网络算法.....	11
5.3.3 支持深度强化学习算法.....	11
5.4 TensorFlow 的应用案例.....	11
5.4.1 Python 语言的编程.....	11
5.4.2 Python 实现 TensorFlow	11

5.4.3 具体的案例.....	11
6 大数据机器学习过程.....	11
6.1 需求的收集、整理.....	11
6.1.1 需求的收集方.....	11
6.1.2 需求的表述方法.....	12
6.2 数据的收集.....	12
6.2.1 能收集哪些数据.....	12
6.2.2 数据的维度.....	12
6.2.3 数据的质量.....	12
6.2.4 数据的管理.....	12
6.3 分析方法的选择.....	12
6.3.1 统计分析方法.....	12
6.3.2 机器学习方法.....	12
6.4 应用开发原型.....	12
6.4.1 原型系统的开发.....	12
6.4.2 原型系统的评估.....	12
6.5 应用的上线.....	12
6.5.1 谁负责使用.....	12
6.5.2 谁负责验收.....	12
6.6 快速迭代开发.....	12
6.6.1 如何快速迭代.....	13
6.6.2 迭代后的上线.....	13
6.7 上线后评估.....	13
6.7.1 客户满意度.....	13
6.7.2 效率提升.....	13
6.7.3 效益评估等.....	13
6.8 效益的评估方法.....	13
6.8.1 经济效益.....	13
6.8.2 社会效益.....	13
6.8.3 管理效益等.....	13
7 大数据机器学习案例.....	13
7.1 【案例 1】客户离网分析案例.....	13
7.2 【案例 2】外来工群体细分及营销.....	13
7.3 【案例 3】人工智能用于网络分析的案例.....	13
7.4 【案例 4】深度学习应用于客户服务系统案例.....	13
8 【附】大数据机器学习应用的方法论——互联网思维.....	14
8.1 互联网思维概述.....	14
8.2 用户思维.....	14
8.2.1 用户是谁.....	14
8.2.2 用户需要什么?.....	14
8.3 极致思维.....	14
8.3.1 产品极致设计.....	14
8.3.2 服务的极致.....	14
8.4 简约思维.....	14

8.4.1	产品的简约	14
8.4.2	简约的流程	14
8.5	迭代思维	14
8.5.1	产品的快速迭代	14
8.5.2	开发过程的快速迭代	14
8.6	流量思维	14
8.6.1	先有客户再有钱	15
8.6.2	扩展客户使用流量	15
8.7	平台思维	15
8.7.1	APP store 的借鉴	15
8.7.2	生态圈的构筑	15
8.8	跨界思维	15
8.8.1	数据的跨界	15
8.8.2	应用的跨界	15
8.9	数据思维	15
8.9.1	收集数据	15
8.9.2	分析数据	15
8.10	社会化思维	15
8.10.1	影响社会	15
8.10.2	社会角度思考	15
8.11	互联网思维在大数据系统中的落地	15
9	总结	16

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

1.6.1 为什么要建生态圈

1.6.2 生态圈的构建方法

2 大数据应用开发方法

2.1 深度学习的方法论有何异同？

2.2 CRISP-DM (Cross-Industry Standard Process for Data Mining) 方法论

2.3 业务理解 (Business Understanding)

2.4 数据理解 (Data Understanding)

2.5 数据准备 (Data Preparation)

2.6 建模 (Modeling)

2.7 评估 (Evaluation)

2.8 部署 (Deployment)

2.9 数据挖掘之外的方法论

3 机器学习基础

3.1 多维分析方法

3.1.1 OLAP 分析

3.1.2 上钻和下钻

3.1.3 用 OLAP 分析问题

3.2 分析算法

3.2.1 回归算法

线性回归

逻辑回归

3.2.2 决策树算法

C4.5 算法

CART 算法

3.2.3 贝叶斯算

法 朴素贝叶斯算法

BBN (Bayesian Belief Network) 算法

3.2.4 基于核的算法

支持向量机 **SVM** 算法

线性判别分析(**Linear Discriminate Analysis** , **LDA**)

3.2.5 聚类算法

K-MEANS 算法

期望最大化算法(**Expectation Maximization** , **ME**)

3.2.6 关联规则算

法 **Apriori** 算法

3.2.7 降低维度算法

主成份分析(**Principle Component Analysis** , **PCA**)算法

偏最小二乘回归(**Partial Least Square Regression** , **PLS**)
算法

3.2.8 集成算法

随机森林算法

梯度推进机

3.3 机器学习方法在电信行业应用举例

4 深度学习基础

4.1 神经网络算法的演进历史

4.1.1 传统的神经网络算法

4.1.2 神经网络算法的特点

4.1.3 神经网络算法的瓶颈点

4.2 神经网络的算法原理

4.2.1 神经元的原理

4.2.2 神经网络的原理

4.2.3 神经网络的 **BP 算法**

4.3 多层神经网络算法

4.3.1 单层神经网络

4.3.2 两层神经网络

4.3.3 多层神经网络——深度学习

4.4 深度学习算法

4.4.1 卷积神经网络 CNN

4.4.2 循环神经网络 RNN

4.4.3 深度神经网络 DNN

4.5 深度学习的应用案例

4.5.1 深度学习用于图像识别

4.5.2 深度学习用于语音识别

4.5.3 深度学习用于语言翻译

5 TensorFlow 等框架

5.1 人工智能的学习框架

5.1.1 开源的概念

5.1.2 学习框架的历程

5.1.3 caffe 等学习框架介绍

5.2 TensorFlow 框架内容

5.2.1 编程模型简介

5.2.2 自编码器介绍

5.2.3 多层感知机介绍

5.3 TensorFlow 支持的算法

5.3.1 支持卷积神经网络算法

5.3.2 支持循环神经网络算法

5.3.3 支持深度强化学习算法

5.4 TensorFlow 的应用案例

5.4.1 Python 语言的编程

5.4.2 Python 实现 TensorFlow

5.4.3 具体的案例

6 大数据机器学习过程

6.1 需求的收集、整理

6.1.1 需求的收集方

6.1.2 需求的表述方法

6.2 数据的收集

6.2.1 能收集哪些数据

6.2.2 数据的维度

6.2.3 数据的质量

6.2.4 数据的管理

6.3 分析方法的选择

6.3.1 统计分析方法

6.3.2 机器学习方法

6.4 应用开发原型

6.4.1 原型系统的开发

6.4.2 原型系统的评估

6.5 应用的上线

6.5.1 谁负责使用

6.5.2 谁负责验收

6.6 快速迭代开发

6.6.1 如何快速迭代

6.6.2 迭代后的上线

6.7 上线后评估

6.7.1 客户满意度

6.7.2 效率提升

6.7.3 效益评估等

6.8 效益的评估方法

6.8.1 经济效益

6.8.2 社会效益

6.8.3 管理效益等

7 大数据机器学习案例

7.1 【案例 1】客户离网分析案例

7.2 【案例 2】外来工群体细分及营销

7.3 【案例 3】人工智能用于网络分析的案例

7.4 【案例 4】深度学习应用于客户服务系统案例

8 【附】大数据机器学习应用的方法论

—— 互联网思维

8.1 互联网思维概述

8.2 用户思维

8.2.1 用户是谁

8.2.2 用户需要什么？

8.3 极致思维

8.3.1 产品极致设计

8.3.2 服务的极致

8.4 简约思维

8.4.1 产品的简约

8.4.2 简约的流程

8.5 迭代思维

8.5.1 产品的快速迭代

8.5.2 开发过程的快速迭代

8.6 流量思维

8.6.1 先有客户再有钱

8.6.2 扩展客户使用流量

8.7 平台思维

8.7.1 APP store 的借鉴

8.7.2 生态圈的构筑

8.8 跨界思维

8.8.1 数据的跨

界 **8.8.2** 应用

的跨界 **8.9** 数据

思维

8.9.1 收集数

据 **8.9.2** 分

析数据 **8.10**

社会化思维

8.10.1 影响社

会

8.10.2 社会角度思考

8.11 互联网思维在大数据系统中的落地

9 总结