

# 《Hadoop大数据分析 with 挖掘实战》

## 图书基本信息

书名：《Hadoop大数据分析 with 挖掘实战》

13位ISBN编号：9787111522656

出版时间：2015-12-1

作者：张良均,樊哲,赵云龙,李成华

页数：289

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：[www.tushu000.com](http://www.tushu000.com)

# 《Hadoop大数据分析 with 挖掘实战》

## 内容概要

《Hadoop大数据分析 with 挖掘实战》共14章，分三个部分：基础篇、实战篇、高级篇。基础篇介绍了数据挖掘、Hadoop大数据的基本原理，实战篇介绍了一个个真实案例，通过对案例深入浅出的剖析，使读者在不知不觉中通过案例实践获得大数据项目挖掘分析经验，同时快速领悟看似难懂的大数据分析与挖掘理论知识。读者在阅读过程中，应充分利用随书配套的案例建模数据，借助TipDM-HB大数据挖掘建模平台，通过上机实验，以快速理解相关知识与理论。

## 书籍目录

前 言

基 础 篇

第1章 数据挖掘基础2

1.1 某知名连锁餐饮企业的困惑2

1.2 从餐饮服务到数据挖掘3

1.3 数据挖掘的基本任务4

1.4 数据挖掘建模过程4

1.4.1 定义挖掘目标4

1.4.2 数据取样5

1.4.3 数据探索6

1.4.4 数据预处理12

1.4.5 挖掘建模14

1.4.6 模型评价14

1.5 餐饮服务中的大数据应用15

1.6 小结15

第2章 Hadoop基础16

2.1 概述16

2.1.1 Hadoop简介16

2.1.2 Hadoop生态系统17

2.2 安装与配置19

2.3 Hadoop原理26

2.3.1 Hadoop HDFS原理26

2.3.2 Hadoop MapReduce原理27

2.3.3 Hadoop YARN原理28

2.4 动手实践30

2.5 小结33

第3章 Hadoop生态系统：Hive34

3.1 概述34

3.1.1 Hive简介34

3.1.2 Hive安装与配置35

3.2 Hive原理38

3.2.1 Hive架构38

3.2.2 Hive的数据模型40

3.3 动手实践41

3.4 小结45

第4章 Hadoop生态系统：HBase46

4.1 概述46

4.1.1 HBase简介46

4.1.2 HBase安装与配置47

4.2 HBase原理50

4.2.1 HBase架构50

4.2.2 HBase与RDBMS51

4.2.3 HBase访问接口52

4.2.4 HBase数据模型53

4.3 动手实践54

4.4 小结61

第5章 大数据挖掘建模平台62

- 5.1 常用的大数据平台62
- 5.2 TipDM-HB大数据挖掘建模平台63
  - 5.2.1 TipDM-HB大数据挖掘建模平台的功能63
  - 5.2.2 TipDM-HB大数据挖掘建模平台操作流程及实例65
  - 5.2.3 TipDM-HB大数据挖掘建模平台的特点67
- 5.3 小结68
- 第6章 挖掘建模69
  - 6.1 分类与预测69
    - 6.1.1 实现过程69
    - 6.1.2 常用的分类与预测算法70
    - 6.1.3 决策树71
    - 6.1.4 Mahout中Random Forests算法的实现原理75
    - 6.1.5 动手实践79
  - 6.2 聚类分析83
    - 6.2.1 常用聚类分析算法83
    - 6.2.2 K-Means聚类算法84
    - 6.2.3 Mahout中K-Means算法的实现原理88
    - 6.2.4 动手实践90
  - 6.3 关联规则93
    - 6.3.1 常用的关联规则算法93
    - 6.3.2 FP-Growth关联规则算法94
    - 6.3.3 Mahout中Parallel Frequent Pattern Mining算法的实现原理98
    - 6.3.4 动手实践100
  - 6.4 协同过滤102
    - 6.4.1 常用的协同过滤算法102
    - 6.4.2 基于项目的协同过滤算法简介102
    - 6.4.3 Mahout中Itembased Collaborative Filtering算法的实现原理103
    - 6.4.4 动手实践106
  - 6.5 小结109
- 实 战 篇
- 第7章 法律咨询数据分析与服务推荐112
  - 7.1 背景与挖掘目标112
  - 7.2 分析方法与过程114
    - 7.2.1 数据抽取120
    - 7.2.2 数据探索分析120
    - 7.2.3 数据预处理125
    - 7.2.4 模型构建130
  - 7.3 上机实验139
  - 7.4 拓展思考140
  - 7.5 小结145
- 第8章 电商产品评论数据情感分析146
  - 8.1 背景与挖掘目标146
  - 8.2 分析方法与过程146
    - 8.2.1 评论数据采集147
    - 8.2.2 评论预处理150
    - 8.2.3 文本评论分词155
    - 8.2.4 构建模型155
  - 8.3 上机实验167
  - 8.4 拓展思考168

## 8.5 小结169

## 第9章 航空公司客户价值分析170

### 9.1 背景与挖掘目标170

### 9.2 分析方法与过程171

#### 9.2.1 数据抽取174

#### 9.2.2 数据探索分析174

#### 9.2.3 数据预处理175

#### 9.2.4 模型构建177

### 9.3 上机实验182

### 9.4 拓展思考183

### 9.5 小结183

## 第10章 基站定位数据商圈分析184

### 10.1 背景与挖掘目标184

### 10.2 分析方法与过程186

#### 10.2.1 数据抽取186

#### 10.2.2 数据探索分析187

#### 10.2.3 数据预处理188

#### 10.2.4 构建模型191

### 10.3 上机实验194

### 10.4 拓展思考195

### 10.5 小结195

## 第11章 互联网电影智能推荐196

### 11.1 背景与挖掘目标196

### 11.2 分析方法与过程197

#### 11.2.1 数据抽取199

#### 11.2.2 构建模型199

### 11.3 上机实验201

### 11.4 拓展思考202

### 11.5 小结203

## 第12章 家电故障备件储备预测分析204

### 12.1 背景与挖掘目标204

### 12.2 分析方法与过程206

#### 12.2.1 数据探索分析207

#### 12.2.2 数据预处理209

#### 12.2.3 构建模型212

### 12.3 上机实验216

### 12.4 拓展思考217

### 12.5 小结217

## 第13章 市供水混凝投药量控制分析218

### 13.1 背景与挖掘目标218

### 13.2 分析方法与过程220

#### 13.2.1 数据抽取221

#### 13.2.2 数据探索分析221

#### 13.2.3 数据预处理223

#### 13.2.4 构建模型227

### 13.3 上机实验237

### 13.4 拓展思考238

### 13.5 小结239

## 第14章 基于图像处理的车辆压双黄线检测240

- 14.1 背景与挖掘目标240
- 14.2 分析方法与过程241
  - 14.2.1 数据抽取242
  - 14.2.2 数据探索分析242
  - 14.2.3 数据预处理242
  - 14.2.4 构建模型249
- 14.3 上机实验250
- 14.4 拓展思考250
- 14.5 小结251
- 高级篇
- 第15章 基于Mahout的大数据挖掘开发254
  - 15.1 概述254
  - 15.2 环境配置255
  - 15.3 基于Mahout算法接口的二次开发258
    - 15.3.1 Mahout算法实例258
    - 15.3.2 Mahout算法接口的二次开发示例259
  - 15.4 小结271
- 第16章 基于TipDM-HB的大数据挖掘二次开发272
  - 16.1 概述272
    - 16.1.1 TipDM-HB大数据挖掘建模平台服务接口272
    - 16.1.2 Apache CXF简介276
  - 16.2 TipDM-HB大数据挖掘建模平台服务开发实例277
    - 16.2.1 环境配置277
    - 16.2.2 开发实例280
  - 16.3 小结288
- 参考资料289

# 《Hadoop大数据分析 with 挖掘实战》

## 精彩短评

- 1、大致操作了一下，了解了一下基本原理，还没有深入的学习，之后需要的时候再细看。
- 2、讲得很浅，不过也算实实在在的案例。

# 《Hadoop大数据分析 with 挖掘实战》

## 版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：[www.tushu000.com](http://www.tushu000.com)