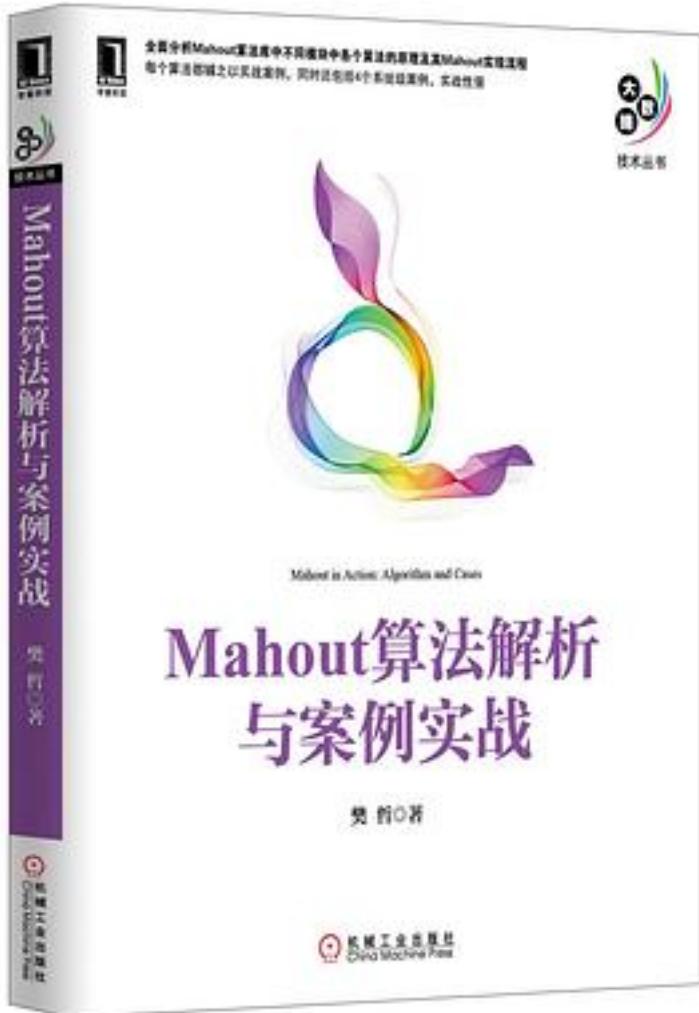


Mahout算法解析与案例实战



[Mahout算法解析与案例实战 下载链接1](#)

著者:樊哲

出版者:机械工业出版社

出版时间:2014-6

装帧:

isbn:9787111467977

【编辑推荐】

全面分析Mahout算法库中不同模块中各个算法的原理及其Mahout实现流程

每个算法都辅之以实战案例，同时还包括4个系统级案例，实战性强

【内容简介】

本书是一本经典的Mahout著作，原理与实战并重。不仅全面分析了Mahout算法库不同模块中的各个算法的原理及其实现流程，而且每个算法都辅之以实战案例。此外，还包括4个系统级案例，实战性非常强。

全书11章共分为三个部分：第一部分为基础篇（第1~2章），首先介绍了Mahout的应用背景、Mahout算法库收录的算法、Mahout的应用实例，以及开发环境的搭建；第二部分为算法篇（第3~7章），分析了Mahout算法库中不同模块的各个算法的原理以及Mahout实现流程，同时在各章节含有每个算法的实战，让读者可以自己运行程序，感受程序运行的各个流程；第三部分为实战篇（第8~11章），通过对4个不同系统案例的分析讲解，让读者了解开发完整的云平台系统的各个流程，即需求分析、系统框架选择及构建、系统功能设计和功能开发。

作者介绍：

樊哲

资深软件开发工程师，精通Java相关技术，专注数据挖掘领域，对Hadoop和Mahout等大数据技术有较深入的研究和丰富的实践，目前正从事Mahout算法开发方面的工作。活跃于CSDN和Hadoop技术论坛等社区，荣获“CSDN2013博客之星”头衔。

目录: 第一部分 基础篇

第1章 Mahout简介 2

1.1 Mahout应用背景 2

1.2 Mahout算法库 3

1.2.1 聚类算法 4

1.2.2 分类算法 5

1.2.3 协同过滤算法 6

1.2.4 频繁项集挖掘算法 7

1.3 Mahout应用 7

1.4 本章小结 8

第2章 Mahout安装配置 9

2.1 Mahout安装前的准备 9

2.1.1 安装JDK 10

2.1.2 安装Hadoop 12

2.2 两种安装方式 20

2.2.1 使用Maven安装 20

2.2.2 下载发布版安装 22

2.3 测试安装 22

2.4 本章小结 24

第二部分 算法篇

第3章 聚类算法 26

3.1 Canopy算法 26

3.1.1 Canopy算法简介 26

3.1.2 Mahout中Canopy算法实现原理 28

3.1.3 Mahout的Canopy算法实战 29

3.1.4 Canopy算法小结 37

3.2 K-Means算法 37

3.2.1 K-Means算法简介 37
3.2.2 Mahout中K-Means算法实现原理 38
3.2.3 Mahout的K-Means算法实战 39
3.2.4 K-Means算法小结 46
3.3 Mean Shift算法 46
3.3.1 Mean Shift算法简介 46
3.3.2 Mahout中Mean Shift算法实现原理 46
3.3.3 Mahout的Mean Shift算法实战 48
3.3.4 Mean Shift算法小结 51

3.4 本章小结 51

第4章 分类算法 52

4.1 Bayesian算法 53

4.1.1 Bayesian算法简介 53

4.1.2 Mahout中Bayesian算法实现原理 55

4.1.3 Mahout的Bayesian算法实战 59

4.1.4 拓展 70

4.1.5 Bayesian算法小结 70

4.2 Random Forests算法 70

4.2.1 Random Forests算法简介 70

4.2.2 Mahout中Random Forests算法实现原理 72

4.2.3 Mahout的Random Forests算法实战 77

4.2.4 拓展 81

4.2.5 Random Forests算法小结 82

4.3 本章小结 83

第5章 协同过滤算法 84

5.1 Distributed Item-Based Collaborative Filtering算法 85

5.1.1 Distributed Item-Based Collaborative Filtering算法简介 85

5.1.2 Mahout中Distributed Item-Based Collaborative Filtering算法实现原理 86

5.1.3 Mahout的Distributed Item-Based Collaborative Filtering算法实战 90

5.1.4 拓展 93

5.1.5 Distributed Item-Based Collaborative Filtering算法小结 94

5.2 Collaborative Filtering with ALSWR算法 94

5.2.1 Collaborative Filtering with ALSWR算法简介 94

5.2.2 Mahout中Collaborative Filtering with ALS-WR算法实现原理 98

5.2.3 Mahout的Collaborative Filtering with ALS-WR算法实战 99

5.2.4 拓展 107

5.2.5 Collaborative Filtering with ALSWR算法小结 107

5.3 本章小结 107

第6章 模式挖掘算法 108

6.1 FP树关联规则算法 109

6.1.1 FP树关联规则算法简介 109

6.1.2 Mahout中Parallel Frequent Pattern Mining算法实现原理 113

6.1.3 Mahout的Parallel Frequent Pattern Mining算法实战 120

6.1.4 拓展 125

6.2 本章小结 126

第7章 Mahout中的其他算法 127

7.1 Dimension Reduction算法 128

7.1.1 Dimension Reduction算法简介 128

7.1.2 Mahout中Dimension Reduction算法实现原理 129

7.1.3 Mahout的Dimension Reduction算法实战 133

7.1.4 拓展 139

7.2 本章小结 142

第三部分 实战篇

第8章 Friend Find系统 144

8.1 系统功能 145
8.1.1 系统管理员 145
8.1.2 普通用户 146
8.1.3 总体功能 146
8.2 数据库设计 147
8.2.1 原始用户数据表 148
8.2.2 注册用户数据表 149
8.2.3 系统管理员表 149
8.2.4 聚类中心表 149
8.3 系统技术框架 150
8.4 系统流程 152
8.4.1 登录 152
8.4.2 注册 153
8.4.3 上传数据 154
8.4.4 调用K-Means算法 155
8.4.5 查看用户分组 157
8.4.6 查看分组情况 158
8.4.7 查看分组成员 159
8.5 系统实现 159
8.5.1 登录 159
8.5.2 注册 161
8.5.3 上传数据 162
8.5.4 调用K-Means算法 163
8.5.5 查看用户分组 167
8.5.6 查看分组情况 167
8.5.7 查看分组成员 168
8.6 本章小结 170
第9章 Wine Identification系统 171
9.1 系统功能 172
9.1.1 用户管理模块 173
9.1.2 随机森林模型建立模块 173
9.1.3 随机森林模型预测模块 173
9.2 系统框架 173
9.3 数据库设计 180
9.3.1 用户表 180
9.3.2 系统常量表 181
9.4 系统流程 181
9.4.1 登录 182
9.4.2 注销 182
9.4.3 权限修改 182
9.4.4 密码修改 183
9.4.5 用户列表 183
9.4.6 数据上传 184
9.4.7 随机森林模型建立 185
9.4.8 随机森林模型评估 186
9.4.9 随机森林模型预测 187
9.5 系统实现 188
9.5.1 登录 188
9.5.2 注销 188
9.5.3 权限修改 189
9.5.4 密码修改 190
9.5.5 用户列表 191
9.5.6 数据上传 193
9.5.7 随机森林模型建立 194

9.5.8 随机森林模型评估	194
9.5.9 随机森林模型预测	195
9.6 本章小结	196
第10章 Dating Recommender系统	197
10.1 系统功能	198
10.1.1 系统管理员功能	198
10.1.2 普通用户功能	199
10.1.3 功能总述	199
10.2 系统框架	200
10.3 数据库设计	203
10.3.1 系统管理员表	203
10.3.2 原始用户推荐信息表	204
10.3.3 基础数据top10表	204
10.4 系统流程	204
10.4.1 登录	205
10.4.2 上传数据	205
10.4.3 推荐分析	206
10.4.4 单用户推荐	210
10.4.5 新用户推荐	211
10.5 算法设计	214
10.5.1 协同过滤算法接口设计	214
10.5.2 top10算法设计	215
10.5.3 新用户推荐算法设计	221
10.6 系统实现	228
10.6.1 登录	228
10.6.2 上传数据	229
10.6.3 推荐分析	230
10.6.4 单用户推荐	232
10.6.5 新用户推荐	234
10.7 本章小结	235
第11章 博客推荐系统	237
11.1 系统功能	238
11.1.1 用户管理	238
11.1.2 建立知识库	239
11.1.3 博客管理	239
11.2 系统框架	240
11.3 数据库设计	246
11.3.1 用户信息表	246
11.3.2 知识库信息表	247
11.3.3 系统常量表	248
11.4 系统流程	248
11.4.1 登录	248
11.4.2 注册	248
11.4.3 密码修改	249
11.4.4 订阅博客查看	249
11.4.5 博客订阅与退订	249
11.4.6 博客推荐	250
11.4.7 上传数据	252
11.4.8 调用FP树关联规则算法	253
11.5 算法设计	260
11.6 系统实现	262
11.6.1 登录	262
11.6.2 注册	263
11.6.3 密码修改	264

11.6.4 订阅博客查看	265
11.6.5 运行FP云算法	266
11.6.6 博客订阅与退订	267
11.6.7 博客推荐	268
11.7 本章小结	270
• • • • •	(收起)

[Mahout算法解析与案例实战](#) [下载链接1](#)

标签

Mahout

数据挖掘

计算机

机器学习

大数据

Mining

评论

太水了。

前边对于算法的介绍还比较好。但是后面实战部分太差了。
调用MAHOUT与HADOOP的平台的接口什么的都没讲。

差评是因为感觉作者不认真连个源码都不给放，学起来很费劲

[Mahout算法解析与案例实战 下载链接1](#)

书评

基本就是骗钱的书，太垃圾了。大家不要买了，这作者人品有问题。还不如本科的毕业论文。还那么贵。真的不值得，浪费时间和金钱，这作者在哪家公司混啊。
基本就是骗钱的书，太垃圾了。大家不要买了，这作者人品有问题。还不如本科的毕业论文。还那么贵。真的不值得，浪费时间和...

[Mahout算法解析与案例实战 下载链接1](#)