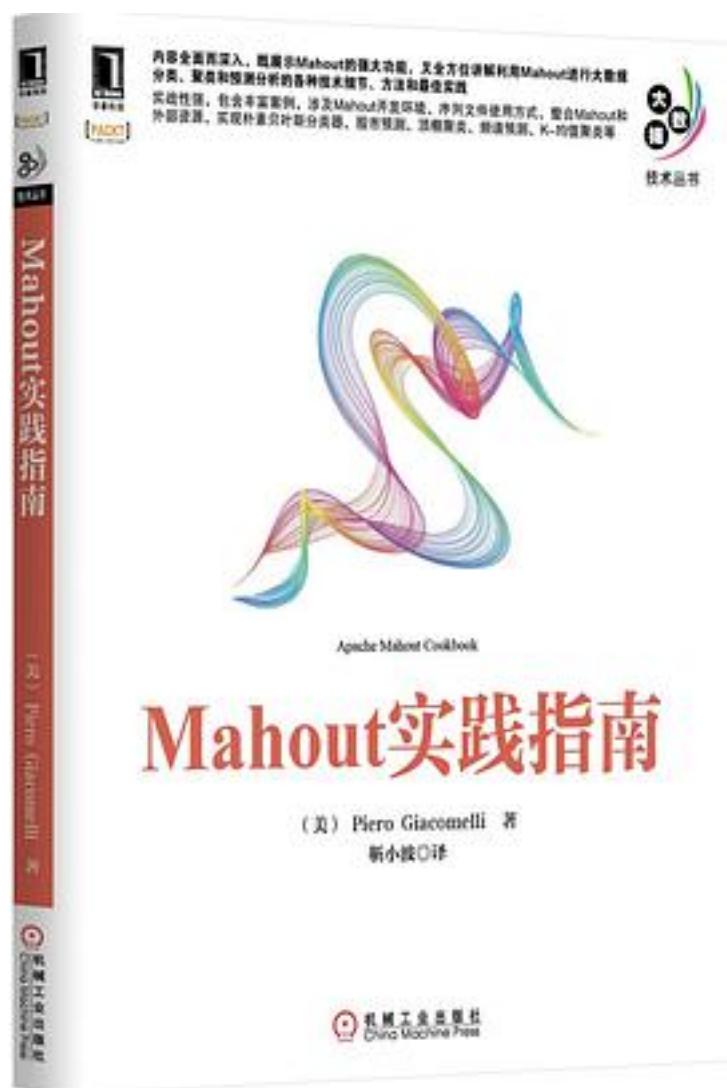


Mahout实践指南（大数据技术丛书）



[Mahout实践指南（大数据技术丛书） 下载链接1](#)

著者: (美) Piero Giacomelli

出版者: 机械工业出版社

出版时间: 2014-6

装帧: 平装

isbn: 9787111467144

【编辑推荐】

内容全面而深入，既展示Mahout的强大功能，又全方位讲解利用Mahout进行大数据分类、聚类 and 预测分析的各种技术细节、方法和最佳实践

实战性强，包含丰富案例，涉及Mahout开发环境、序列文件使用方式、整合Mahout和外部资源、实现朴素贝叶斯分类器、股市预测、顶棚聚类、频谱预测、K-均值聚类等

【内容简介】

本书是软件开发专家数十年行业经验的结晶，深入浅出地论述如何使用Mahout进行数据分类、聚类 and 预测，涉及Mahout开发环境、序列文件使用方式、整合Mahout和外部资源、实现朴素贝叶斯分类器、股市预测、顶棚聚类、频谱预测、K-均值聚类等。本书是面向编程的，不涉及深奥的理论，简单、易学，可以帮助读者快速掌握Mahout的基本用法，实用性很强。

全书共分10章。第1章介绍如何在单台机器上创建完整的Mahout开发环境。第2章重点介绍序列文件的使用方式。第3章详细介绍如何使用命令行工具和代码从RDBMS中读写数据。第4章详细讨论朴素贝叶斯分类器和互补朴素贝叶斯分类器的使用方法。第5章介绍如何使用logistic回归和随机森林预测股市。第6章描述Mahout框架中最常用的算法，包括大数据的聚类分析和分类。第7章描述频谱聚类的使用方式。第8章描述使用K-均值（包括序列方式和MapReduce方式）对主题中的文本文档进行分类。第9章介绍频繁模式挖掘算法的使用方式。第10章描述使用遗传算法解决旅行商问题和提取规则。

作者介绍:

Piero Giacomelli

资深软件技术专家，精通Java、.NET和PHP等多种编程语言，尤其对Java语言有独到见解。他曾先后在多家大中型公司担任行政和技术职务，包括航空航天、网络服务、塑料制造业和电子健康协会。他在FP7欧盟项目中参与了多项欧盟研究基金资助的工程，如CHRONIOUS、I-DONT-FALL、FEARLESS和CHROMED等。他在科研期刊上发表了若干篇论文，并两次获得由IARIA颁发的最佳论文奖。2012年，他出版了Apache HornetQ框架标准参考书《HornetQ Messaging Developer's Guide》。

目录: 第1章 Mahout入门 / 1

秘笈1 安装Java和Hadoop / 1

秘笈2 设置Maven和NetBeans开发环境 / 6

秘笈3 编写一个基本的推荐系统 / 9

第2章 使用序列文件——什么时候和为什么 / 19

秘笈4 从命令行创建序列文件 / 20

秘笈5 编写代码创建序列文件 / 23

秘笈6 编码实现读取序列文件 / 28

第3章 将Mahout和外部资源整合 / 33

秘笈7 导入外部资源到HDFS / 34

秘笈8 将数据从HDFS导入到RDBMS / 43

秘笈9 创建一个Sqoop作业来处理RDBMS / 45

秘笈10 使用Sqoop API导入数据 / 47

第4章 实现朴素贝叶斯分类器 / 49

秘笈11 使用Mahout文本分类器演示基本的使用样例 / 50

秘笈12 编码实现朴素贝叶斯分类器 / 60

秘笈13 通过命令行使用互补朴素贝叶斯 / 64

秘笈14 编码使用互补朴素贝叶斯分类器 / 65

第5章 股市预测 / 67
秘笈15 为logistic回归准备数据 / 67
秘笈16 使用logistic预测GOOG股票动态 / 71
秘笈17 通过Java编码使用自适应的logistic回归 / 76
秘笈18 在大规模的数据集上使用logistic回归 / 79
秘笈19 使用随机森林预测市场动态 / 83
第6章 顶棚聚类 / 87
秘笈20 基于命令行的顶棚聚类 / 87
秘笈21 基于带参数命令行的顶棚聚类 / 91
秘笈22 通过Java代码使用顶棚聚类 / 95
秘笈23 编写你自己的距离估计 / 98
第7章 频谱聚类 / 101
秘笈24 通过命令行使用EigenCuts / 101
秘笈25 在Java代码中使用EigenCuts / 104
秘笈26 从原始数据创建相似度矩阵 / 108
秘笈27 使用频谱聚类进行图像分割 / 114
第8章 K-均值聚类 / 119
秘笈28 在Java代码中使用K-均值聚类 / 119
秘笈29 使用K-均值聚类对交通事故进行聚类 / 124
秘笈30 使用MapReduce进行K-均值聚类 / 128
秘笈31 命令行方式使用K-均值聚类 / 132
第9章 软计算 / 139
秘笈32 使用Mahout进行频繁模式挖掘 / 139
秘笈33 为频繁模式挖掘创建评价准则 / 142
秘笈34 在Java代码中使用频繁模式挖掘 / 147
秘笈35 使用LDA创建主题 / 153
第10章 实现遗传算法 / 159
秘笈36 设置Mahout以便使用遗传算法 / 159
秘笈37 在图上使用遗传算法 / 163
秘笈38 在Java代码中使用遗传算法 / 167
· · · · · (收起)

[Mahout实践指南（大数据技术丛书）_下载链接1](#)

标签

机器学习

Mahout

计算机

软件开发

推荐算法

评论

入门读物

讲Mahout，居然没有推荐？

原理性的东西比较少，基本上是实践，内容较浅显，适合入门。

太薄了。并且介绍的也不够详细。实现的算法也比较少 不过还是不错的。
MAHOUT的教程太少了

[Mahout实践指南（大数据技术丛书） 下载链接1](#)

书评

[Mahout实践指南（大数据技术丛书） 下载链接1](#)