

大数据架构和算法实现之路：电商系统的技术实战



[大数据架构和算法实现之路：电商系统的技术实战 下载链接1](#)

著者:黄申

出版者:机械工业出版社

出版时间:2017-6

装帧:平装

isbn:9787111569695

本书介绍了一些主流技术在商业项目中的应用，包括机器学习中的分类、聚类和线性回归，搜索引擎，推荐系统，用户行为跟踪，架构设计的基本理念及常用的消息和缓存机制。在这个过程中，我们有机会实践R、Mahout、Solr、Elasticsearch、Hadoop、HBase、Hive、Flume、Kafka、Storm等系统。如前所述，本书最大的特色就是，从商业需

求出发演变到合理的技术方案和实现，因此根据不同的应用场景、不同的数据集合、不同的进阶难度，我们为读者提供了反复温习和加深印象的机会。

作者介绍:

目录: 目录

推荐序

前言

引子 1

第一篇 支持高效的运营

第1章 方案设计和选型：分类 5

1.1 分类的基本概念 6

1.2 分类任务的流程 7

1.3 算法：朴素贝叶斯和K最近邻 8

1.3.1 朴素贝叶斯 8

1.3.2 K最近邻 9

1.4 分类效果评估 10

1.5 相关软件：R和Mahout 12

1.5.1 R简介 12

1.5.2 Mahout简介 13

1.5.3 Hadoop简介 14

1.6 案例实践 17

1.6.1 实验环境设置 17

1.6.2 中文分词 18

1.6.3 使用R进行朴素贝叶斯分类 22

1.6.4 使用R进行K最近邻分类 37

1.6.5 单机环境使用Mahout运行朴素贝叶斯分类 39

1.6.6 多机环境使用Mahout运行朴素贝叶斯分类 47

1.7 更多的思考 58

第2章 方案设计和选型：聚类 60

2.1 聚类的基本概念 60

2.2 算法：K均值和层次型聚类 61

2.2.1 K均值聚类 61

2.2.2 层次型聚类 62

2.3 聚类效果评估 64

2.4 案例实践 66

2.4.1 使用R进行K均值聚类 66

2.4.2 使用Mahout进行K均值聚类 69

第3章 方案设计和选型：因变量连续的回归分析 74

3.1 线性回归的基本概念 74

3.2 案例实践 76

3.2.1 实验环境设置 76

3.2.2 R中数据的标准化 78

3.2.3 使用R的线性回归分析 81

第二篇 为顾客发现喜欢的商品：

基础篇

第4章 方案设计和选型：搜索 94

4.1 搜索引擎的基本概念 94

4.1.1 相关性 95

4.1.2 及时性 97

4.2 搜索引擎的评估 100

4.3 为什么不是数据库 103

4.4 系统框架	104
4.4.1 离线预处理	104
4.4.2 在线查询	107
4.5 常见的搜索引擎实现	108
4.5.1 Lucene简介	108
4.5.2 Solr简介	113
4.5.3 Elasticsearch简介	120
4.6 案例实践	123
4.6.1 实验环境设置	123
4.6.2 基于Solr的实现	123
4.6.3 基于Elasticsearch的实现	154
4.6.4 统一的搜索API	175
第三篇 为顾客发现喜欢的商品：高级篇	
第5章 方案设计和选型：NoSQL和搜索的整合	195
5.1 问题分析	195
5.2 HBase简介	196
5.3 结合HBase和搜索引擎	203
5.4 案例实践	204
5.4.1 实验环境设置	204
5.4.2 HBase的部署	205
5.4.3 HBase和搜索引擎的集成	211
第6章 方案设计和选型：查询分类和搜索的整合	219
6.1 问题分析	219
6.2 结合分类器和搜索引擎	219
6.3 案例实践	225
6.3.1 实验环境设置	225
6.3.2 构建查询分类器	226
6.3.3 定制化的搜索排序	229
6.3.4 整合查询分类和定制化排序	236
第7章 方案设计和选型：个性化搜索	245
7.1 问题分析	245
7.2 结合用户画像和搜索引擎	245
7.3 案例实践	249
7.3.1 用户画像的读取	250
7.3.2 个性化搜索引擎	253
7.3.3 结果对比	260
第8章 方案设计和选型：搜索分片	267
8.1 问题分析	267
8.2 利用搜索的分片机制	269
8.3 案例实践	271
8.3.1 Solr路由的实现	271
8.3.2 Elasticsearch路由的实现	278
第9章 方案设计和选型：搜索提示	283
9.1 问题分析	283
9.2 案例实践：基础方案	284
9.2.1 Solr搜索建议和拼写纠错的实现	284
9.2.2 Elasticsearch搜索建议和拼写纠错的实现	286
9.3 改进方案	291
9.4 案例实践：改进方案	294
第10章 方案设计和选型：推荐	303
10.1 推荐系统的基本概念	305
10.2 推荐的核心要素	306
10.2.1 系统角色	306
10.2.2 相似度	307

10.2.3 相似度传播框架	307
10.3 推荐系统的分类	307
10.4 混合模型	311
10.5 系统架构	312
10.6 Mahout中的推荐算法	313
10.7 电商常见的推荐系统方案	314
10.7.1 电商常见的推荐系统方案	314
10.7.2 相似度的计算	317
10.7.3 协同过滤	319
10.7.4 结果的查询	320
10.8 案例实践	321
10.8.1 基于内容特征的推荐	321
10.8.2 基于行为特征的推荐	341
第四篇 获取数据，跟踪效果	
第11章 方案设计和选型：行为跟踪	369
11.1 基本概念	370
11.1.1 网站的核心框架	370
11.1.2 行为数据的类型	371
11.1.3 行为数据的模式	372
11.1.4 设计理念	374
11.2 使用谷歌分析	375
11.3 自行设计之Flume、HDFS和Hive的整合	378
11.3.1 数据的收集——Flume简介	378
11.3.2 数据的存储——Hadoop HDFS回顾	382
11.3.3 批量数据分析——Hive简介	383
11.3.4 Flume、HDFS和Hive的整合方案	386
11.4 自行设计之Flume、Kafka和Storm的整合	386
11.4.1 实时性数据分析之Kafka简介	386
11.4.2 实时性数据分析之Storm简介	388
11.4.3 Flume、Kafka和Storm的整合方案	390
11.5 案例实践	391
11.5.1 数据模式的设计	392
11.5.2 实验环境设置	392
11.5.3 谷歌分析实战	394
11.5.4 自主设计实战之Flume、HDFS和Hive的整合	401
11.5.5 自主设计实战之Flume、Kafka和Storm的整合	410
11.6 更多的思考	424
后记	425
• • • • •	(收起)

[大数据架构和算法实现之路：电商系统的技术实战 下载链接1](#)

标签

大数据

架构

数据

计算机

编程

电商算法

图书馆k

图书馆

评论

实用性超强

本书介绍了一些主流技术在商业项目中的应用，包括机器学习中的分类、聚类和线性回归，搜索引擎，推荐系统，用户行为跟踪，架构设计的基本理念及常用的消息和缓存机制。在这个过程中，我们有机会实践R、Mahout、Solr、Elasticsearch、Hadoop、HBase、Hive、Flume、Kafka、Storm等系统。

作为续作，本书最大的特色是加入了大量的实践和代码，非常实用

很不错，我从国图借的，开阔视野，但是不适合新手读书，业务背景要求太高了！准备过几年再读（到时候，推荐会变吧）

[大数据架构和算法实现之路：电商系统的技术实战 下载链接1](#)

[大数据架构和算法实现之路：电商系统的技术实战 下载链接1](#)