

《大数据是这样计算的：XLab实例入门》

图书基本信息

书名：《大数据是这样计算的：XLab实例入门》

13位ISBN编号：9787121282313

出版时间：2016-4

作者：杨旭

页数：340

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.tushu111.com

《大数据是这样计算的：XLab实例入门》

内容概要

随着大数据分析方法的不断发展、分析工具的改进，大数据分析的门槛降低了。基于大数据算法平台，数据分析师们已经不再被数据的“大”所困扰，他们可以轻松地探索大数据，挖掘大数据的价值。

《大数据是这样计算的：XLab实例入门》侧重介绍大数据分析方法和算法的应用，选取了10个不同领域的真实数据集，针对每个数据的特点，选择适合的方法和算法，和读者一起体验数据探索、数据分析、建模预测的过程；通过实例计算的结果，读者会更加了解各种方法和算法的长处和局限。《大数据是这样计算的：XLab实例入门》的实例都运行在大数据算法平台XLab上。

《大数据是这样计算的：XLab实例入门》适合对大数据分析感兴趣的读者阅读，也可供数据分析师、算法工程师等专业人员参考使用。

《大数据是这样计算的：XLab实例入门》

作者简介

杨旭，吉林长春人，2004年获南开大学数学博士学位；随后在南开大学信息学院从事博士后研究工作；2006年加入微软亚洲研究院，进行符号计算、大规模矩阵计算及机器学习算法研究；2010年加入阿里巴巴，从事大数据相关的统计和机器学习算法研发。著有《重构大数据统计》。

书籍目录

第1章 入门

1

1.1 打开大数据之门

1

1.2 接触大数据

2

1.3 数据初探索

4

1.4 进一步分析

10

1.5 训练和预测

17

第2章 简介

23

2.1 主界面介绍

23

2.1.1 工具栏

24

2.1.2 数据列表

26

2.1.3 窗口管理器

28

2.1.4 工作界面

28

2.2 数据查看、运行

29

2.2.1 文本显示

31

2.2.2 复制部分数据

31

2.2.3 显示/隐藏列

32

2.3 脚本编辑、运行

33

2.4 如何获得帮助信息

37

2.4.1 查看帮助手册

37

2.4.2 界面上的帮助信息

38

2.4.3 脚本函数的帮助信息

40

2.5 执行SQL语句

44

第3章 全球机场信息

46

3.1 数据可视化

47	
3.2	统计分析
58	
3.3	大中型机场的分析
62	
3.4	海拔高度排行
64	
3.5	数据的关联关系
68	
	第4章 股票价格
73	
4.1	数据处理
74	
4.2	数据探索
78	
4.3	数据展开
81	
4.3.1	组合使用基本函数进行变换
83	
4.3.2	利用专门函数一步到位
84	
4.4	各股趋势比较
85	
4.5	总体趋势
87	
	第5章 标准普尔500指数
91	
5.1	数据类型转换
92	
5.2	各指标间的线性关系
93	
5.3	按时间变化趋势
96	
5.4	数据对比
100	
	第6章 鸢尾花数据集
107	
6.1	属性间的关系
108	
6.2	聚类
110	
6.2.1	K-Means聚类
111	
6.2.2	EM聚类
120	
6.3	二分类数据子集
124	
6.3.1	使用训练、预测窗体
125	

6.3.2 调用训练、预测脚本

131

第7章 MovieLens数据集

136

7.1 数据变换

137

7.2 统计

138

7.3 排行榜

145

7.4 分类排行榜

147

7.5 影片关联分析

149

7.6 属性扩展

157

第8章 汽车评价数据集

161

8.1 数据图示化

162

8.2 对比分析

165

8.3 决策树

169

第9章 Twitter数据

174

9.1 用户信息分析

175

9.2 用户粉丝数量的情况

184

9.3 粉丝的情况

192

9.4 “粉”与“被粉”

197

9.5 信息传播速度

204

9.6 哪些用户更重要

208

9.6.1 粉丝最多的用户

209

9.6.2 用户排名

211

第10章 随机数据

215

10.1 数据生成

215

10.2 计算 值

217

10.3 中心极限定理

222	
第11章 新浪网页数据	
230	
11.1 分词	
232	
11.2 有区分度的单词	
234	
11.3 选择特征	
235	
11.3.1 卡方检验	
236	
11.3.2 信息增益	
238	
11.4 主题模型	
242	
11.4.1 潜在语义分析	
242	
11.4.2 概率潜在语义分析	
253	
11.4.3 LDA模型	
272	
11.5 单词映射为向量	
289	
第12章 2014年阿里巴巴大数据竞赛	
294	
12.1 试题介绍	
294	
12.2 数据	
296	
12.3 思路	
298	
12.3.1 用户和品牌的各种特征	
298	
12.3.2 二分类模型训练	
300	
12.3.3 比赛考核目标	
301	
12.4 计算训练数据集	
302	
12.4.1 原始数据划分	
303	
12.4.2 计算特征	
304	
12.4.3 数据预处理标识	
304	
12.4.4 用户-品牌联合特征	
308	
12.4.5 用户特征	
313	

12.4.6 品牌特征	317
12.4.7 整合训练数据的特征	322
12.4.8 计算标签	323
12.5 二分类模型训练	324
12.5.1 正负样本配比	324
12.5.2 朴素贝叶斯算法	325
12.5.3 逻辑回归算法	326
12.5.4 随机森林算法	327
12.6 提交预测结果	328

《大数据是这样计算的：XLab实例入门》

精彩短评

1、好歹也提供一份数据，提供个xlab的下载链接吧...

《大数据是这样计算的：XLab实例入门》

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.tushu111.com