

《数据科学:R语言实战》

图书基本信息

书名：《数据科学:R语言实战》

13位ISBN编号：9787115435901

出版时间：2016-11-1

作者：丹·图米 (Dan Toomey)

页数：327

译者：刘丽君,李成华,卢青峰

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

《数据科学:R语言实战》

内容概要

本书讲述的是R语言在数据科学中的应用，目标读者是从事不同行业的数据分析师、数据挖掘工程师、机器学习工程师、自然语言处理工程师、数据科学家，以及从事大数据和人工智能领域的工作者、学生、老师等。

本书的优点在于其通俗易懂、容易上手，每一个实例都有现成的数据和源代码，读者不仅能理解整个案例的来龙去脉，还可以直接编译本书提供的所有源代码，从而了解怎么从实际问题转变成可实现的代码，感受R语言的魅力，让数据产生价值。这种学习和实践相结合的方式非常适合初学者和有一定经验的数据分析师。

本书的内容涵盖了基于数据挖掘的常用模型，包括分类、聚类、关联分析、预测、异常检测等，还包括机器学习的常用算法和自然语言处理、数据可视化等内容。本书内容全面，做到了易读、易用、易理解、易实现、易上手，是不可多得的R语言书籍。

作者简介

作者简介：

Dan Toomey具有20多年开发应用程序方面的经验，曾在多个行业及公司担任不同的职位，包括投稿人、副总裁及首席技术官。近10年，Dan一直在美国马萨诸塞州东部地区的公司工作。Dan以Dan Toomey软件公司的名义，成为这些领域的开发承包商。

译者简介：

刘丽君，韩国国立全北大学博士，加拿大圣西维尔大学博士后，一直从事物联网、工业大数据等方面的数据分析、市场分析等工作，目前任武汉泰迪智慧科技有限公司CEO，对数据敏感，并对数据怎么转变成价值、数据与商业的关系有独到见解。

李成华，数据挖掘与机器学习方向博士，约克大学博士后，麻省理工学院访问科学家，曾任海信集团数据挖掘专家，京东深度神经网络实验室首席科学家，长期从事数据挖掘、机器学习、深度学习和自然语言处理等方面的研究和工作，擅长自动问答以及基于自然语言的人机交互。

卢青峰，硕士毕业于美国威斯康辛州立大学，毕业后从事数据分析、挖掘等相关工作至今。曾先后在敦煌网、百度、京东等行业领先的公司从事数据挖掘、用户行为分析、推荐等工作。

书籍目录

第1章 模式的数据挖掘 1

1.1 聚类分析 2

1.1.1 K—means 聚类 3

1.1.2 K—medoids 聚类 7

1.1.3 分层聚类 12

1.1.4 期望最大化 15

1.1.5 密度估计 21

1.2 异常检测 24

1.2.1 显示异常值 25

1.2.2 计算异常 28

1.3 关联规则 30

1.4 问题 33

1.5 总结 34

第2章 序列的数据挖掘 35

2.1 模式 35

2.1.1 Eclat 36

2.1.2 rulesNBMiner 40

2.1.3 Apriori 43

2.1.4 用TraMineR确定序列 47

2.1.5 序列相似点 54

2.2 问题 57

2.3 总结 57

第3章 文本挖掘 59

3.1 功能包 60

3.1.1 文本处理 60

3.1.2 文本集群 69

3.2 问题 80

3.3 总结 80

第4章 数据分析——回归分析 81

4.1 功能包 81

4.1.1 简单回归 81

4.1.2 多次回归 88

4.1.3 多变量回归分析 94

4.1.4 稳健回归 100

4.2 问题 106

4.3 总结 106

第5章 数据分析——相关性 107

5.1 功能包 107

5.1.1 基本相关性 108

5.1.2 可视化相关性 112

5.1.3 协方差 114

5.1.4 皮尔森相关性 117

5.1.5 多分格相关性 118

5.1.6 四分相关性 122

5.1.7 异构相关矩阵 126

5.1.8 部分相关性 128

5.2 问题 129

5.3总结	129
第6章 数据分析——聚类	131
6.1功能包	131
6.2K—means聚类	132
6.2.1示例	132
6.2.2Medoids集群	140
6.2.3cascadeKM函数	142
6.2.4基于贝叶斯定理信息选取集群	144
6.2.5仿射传播聚类	146
6.2.6用于估测集群数量的间隙统计量	149
6.2.7分级聚类	151
6.3问题	153
6.4总结	154
第7章 数据可视化——R图形	155
7.1功能包	155
7.1.1交互式图形	156
7.1.2lattice功能包	160
7.1.3ggplot2功能包	169
7.2问题	180
7.3总结	181
第8章 数据可视化——绘图	183
8.1功能包	183
8.2散点图	183
8.2.1回归线	187
8.2.2lowess线条	188
8.2.3scatterplot函数	189
8.2.4Scatterplot矩阵	192
8.2.5密度散点图	197
8.3直方图和条形图	200
8.3.1条形图	200
8.3.2直方图	203
8.3.3ggplot2	203
8.3.4词云	204
8.4问题	206
8.5总结	206
第9章 数据可视化——三维	207
9.1功能包	207
9.2生成三维图形	208
9.2.1LatticeCloud——三维散点图	212
9.2.2scatterplot3d	215
9.2.3scatter3d	216
9.2.4cloud3d	218
9.2.5RgoogleMaps	220
9.2.6vrmlgenbar3D	221
9.2.7大数据	223
9.2.8研究方向	228
9.3问题	234
9.4总结	234
第10章 机器学习实战	235

- 10.1功能包 235
- 10.2数据集 236
 - 10.2.1数据划分 240
 - 10.2.2模型 241
 - 10.2.3train方法 254
- 10.3问题 264
- 10.4总结 264
- 第11章 用机器学习预测事件 265
 - 11.1自动预测功能包 265
 - 11.1.1时间序列 266
 - 11.1.2SMA函数 272
 - 11.1.3分解函数 273
 - 11.1.4指数平滑法 274
 - 11.1.5预测 277
 - 11.1.6霍尔特指数平滑法 281
 - 11.2问题 293
 - 11.3总结 293
- 第12章 监督学习和无监督学习 295
 - 12.1功能包 296
 - 12.1.1监督学习 296
 - 12.1.2无监督学习 316
 - 12.2问题 327
 - 12.3总结 327

《数据科学:R语言实战》

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.tushu111.com