

# 《R语言与数据挖掘最佳实践和经典案例》

## 图书基本信息

书名：《R语言与数据挖掘最佳实践和经典案例》

13位ISBN编号：9787111475410

出版时间：2014-9

作者：(澳)Yanchang Zhao

译者：陈健,黄琰

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](http://www.tushu111.com)

# 《R语言与数据挖掘最佳实践和经典案例》

## 内容概要

### 【内容简介】

数据挖掘技术已经广泛用于政府机关、银行、保险、零售、电信、医药和研究领域。最近，越来越多的数据挖掘工作开始使用R工具来完成，R是一个用于统计计算和制图的免费软件。在最近的调查中，R已经被评为数据挖掘领域最流行的工具。本书介绍将R语言用于数据挖掘应用（从学术研究到工业应用），从大量数据中提取出有用知识的各种实用方法。

本书面向数据挖掘领域的研究人员、数据挖掘方向的研究生，以及数据挖掘工程师和分析师，对于学习数据挖掘课程的学生来说具有巨大的参考价值，对于参加数据挖掘与分析的行业培训课程的人来说是非常有用的资料。

### 本书主要特色

- 介绍了R用于数据挖掘应用的案例，涵盖了最常用的数据挖掘技术。

- 提供了代码示例和数据，以便读者可以轻松地学习数据挖掘技术。

- 现实应用中的特色案例研究有助于读者将学到的技术应用到自己的工作和研究中。

# 《R语言与数据挖掘最佳实践和经典案例》

## 作者简介

Yanchang Zhao 从2009年起担任澳大利亚政府部门的高级数据挖掘分析师。在加入澳大利亚政府部门之前，他是悉尼科技大学工程和信息学院博士后研究员。他的研究兴趣包括聚类分析、关联规则、时间序列、孤立点检测、数据挖掘应用等，当前关注在数据挖掘应用中使用R语言。他是IEEE高级会员和澳大利亚分析专业人员协会成员。他发表了50多篇数据挖掘研究和应用方面的论文，并独立或与他人合作编写了3本著作。

## 书籍目录

### 第1章 简介1

#### 1.1 数据挖掘1

#### 1.2 R1

#### 1.3 数据集2

##### 1.3.1 iris数据集2

##### 1.3.2 bodyfat数据集3

### 第2章 数据的导入与导出4

#### 2.1 R数据的保存与加载4

#### 2.2 .CSV文件的导入与导出4

#### 2.3 从SAS中导入数据5

#### 2.4 通过ODBC导入与导出数据6

##### 2.4.1 从数据库中读取数据7

##### 2.4.2 从Excel文件中导入与导出数据7

### 第3章 数据探索8

#### 3.1 查看数据8

#### 3.2 探索单个变量10

#### 3.3 探索多个变量12

#### 3.4 更多探索15

#### 3.5 将图表保存到文件中19

### 第4章 决策树与随机森林21

#### 4.1 使用party包构建决策树21

#### 4.2 使用rpart包构建决策树24

#### 4.3 随机森林29

### 第5章 回归分析33

#### 5.1 线性回归33

#### 5.2 逻辑回归38

#### 5.3 广义线性回归38

#### 5.4 非线性回归40

### 第6章 聚类41

#### 6.1 k-means聚类41

#### 6.2 k-medoids聚类43

#### 6.3 层次聚类45

#### 6.4 基于密度的聚类46

### 第7章 离群点检测50

#### 7.1 单变量的离群点检测50

#### 7.2 局部离群点因子检测53

#### 7.3 用聚类方法进行离群点检测56

#### 7.4 时间序列数据的离群点检测58

#### 7.5 讨论59

### 第8章 时间序列分析与挖掘60

#### 8.1 R中的时间序列数据60

#### 8.2 时间序列分解60

#### 8.3 时间序列预测62

#### 8.4 时间序列聚类63

##### 8.4.1 动态时间规整63

##### 8.4.2 合成控制图的时间序列数据64

##### 8.4.3 基于欧氏距离的层次聚类65

- 8.4.4 基于DTW距离的层次聚类66
- 8.5 时间序列分类67
  - 8.5.1 基于原始数据的分类67
  - 8.5.2 基于特征提取的分类68
  - 8.5.3 k-NN分类69
- 8.6 讨论70
- 8.7 延伸阅读70
- 第9章 关联规则71
  - 9.1 关联规则的基本概念71
  - 9.2 Titanic数据集71
  - 9.3 关联规则挖掘73
  - 9.4 消除冗余78
  - 9.5 解释规则79
  - 9.6 关联规则的可视化80
  - 9.7 讨论与延伸阅读82
- 第10章 文本挖掘84
  - 10.1 Twitter的文本检索84
  - 10.2 转换文本85
  - 10.3 提取词干86
  - 10.4 建立词项-文档矩阵88
  - 10.5 频繁词项与关联90
  - 10.6 词云91
  - 10.7 词项聚类92
  - 10.8 推文聚类94
    - 10.8.1 基于k-means算法的推文聚类94
    - 10.8.2 基于k-medoids算法的推文聚类96
  - 10.9 程序包、延伸阅读与讨论98
- 第11章 社交网络分析99
  - 11.1词项网络99
  - 11.2推文网络102
  - 11.3双模式网络107
  - 11.4讨论与延伸阅读110
- 第12章 案例：房价指数的分析与预测111
  - 12.1HPI数据导入111
  - 12.2HPI数据探索112
  - 12.3HPI趋势与季节性成分118
  - 12.4HPI预测120
  - 12.5房地产估价122
  - 12.6讨论122
- 第13章 案例：客户回复预测与效益最大化123
  - 13.1简介123
  - 13.2KDD Cup 1998的数据123
  - 13.3数据探索131
  - 13.4训练决策树137
  - 13.5模型评估140
  - 13.6选择最优决策树143
  - 13.7评分145
  - 13.8讨论与总结148
- 第14章 案例：内存受限的大数据预测模型150

14.1简介	150
14.2研究方法	150
14.3数据与变量	151
14.4随机森林	152
14.5内存问题	153
14.6样本数据的训练模型	154
14.7使用已选变量建立模型	156
14.8评分	162
14.9输出规则	168
14.9.1以文本格式输出规则	168
14.9.2输出SAS规则的得分	172
14.10总结与讨论	177
第15章 在线资源	178
15.1R参考文档	178
15.2R	178
15.3数据挖掘	179
15.4R的数据挖掘	180
15.5R的分类与预测	181
15.6R的时间序列分析	181
15.7R的关联规则挖掘	181
15.8R的空间数据分析	181
15.9R的文本挖掘	182
15.10R的社交网络分析	182
15.11R的数据清洗与转换	182
15.12R的大数据与并行计算	182
R语言数据挖掘参考文档	184
参考资料	197
通用索引	201
包索引	203
函数索引	204

# 《R语言与数据挖掘最佳实践和经典案例》

## 精彩短评

- 1、写的一般般，算法与实现代码都没有给出比较好的解释，对初学者不太友好，只能充当其他数据挖掘类书籍的补充了。。。
- 2、太简单了吧
- 3、通过多个简洁明了的案例，介绍了常用的数据挖掘技术如何在R中实现，阅读学习本书时，需要准备一些原理、算法及思路知识，多用help。
- 4、本书和《数据挖掘与R语言》有些类似，唯一印象深刻的是对随机森林模型的讲解。
- 5、挺好懂的一本书，也比较容易上手。
- 6、例子也太简单了
- 7、挺好的一本书
- 8、代码挺多的。

# 《R语言与数据挖掘最佳实践和经典案例》

## 精彩书评

1、2015最新数据挖掘入门到精通—R语言视频教程课程观看地址

： <http://www.xuetuwuyou.com/course/59>课程出自学途无忧网： <http://www.xuetuwuyou.com/>课程介绍一、课程所用软件：R 3.2.2 (64位) RStudio二、课程涉及到的技术点：1) R语言的基本语法、函数2) R中实用性很强的包3) 模式识别、分类预测算法原理及其实现三、课程学习目标：本课程讲解理论的同时结合大量的案例，让学习者可以快速掌握数据挖掘技能，并利用R数据处理、画图、实现数据挖掘模型的建立。学习完本课程，学习者能达到以下目标：1) 掌握基本R用法；2) 用R进行描述性统计分析、进行数据处理和数据可视化；3) 缺失值的清洗能力；4) 用R语言建立数据挖掘模型；四、课程大纲：第一章：基本概念介绍第1课、数据挖掘、R语言概念介绍第2课、软件安装和数据的读、写、修改第3课、基本概念讲解（向量、矩阵、因子、数据框、列表）第4课、基本图形的讲解和绘制第二章：实用软件包介绍及应用第5课、plyr包主函数讲解第6课、plyr包辅助函数讲解第7课、Ggpolt2介绍第8课、Ggpolt2实践第9课、reshape2包的讲解和实际操作第10课、课缺失值的处理第三章：算法讲解及应用第11课、knn原理简介第12课、knn算法实际操作第13课、决策树的理论讲解第14课、决策树实操第15课、人工神经网络的介绍1第16课、人工神经网络介绍2第17课、人工神经网络实操1第18课、人工神经网络实操2第19课、支持向量机原理介绍第20课、支持向量机的实操

# 《R语言与数据挖掘最佳实践和经典案例》

## 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](http://www.tushu111.com)