

# 《R语言与网站分析》

## 图书基本信息

书名：《R语言与网站分析》

13位ISBN编号：9787111459717

出版时间：2014-4

作者：李明

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语言与网站分析》

## 内容概要

《R语言与网站分析》从互联网应用角度对R语言如何实现数据挖掘和指标分析等问题做了阐述。通过诸多真实应用案例的分析，作者试图为读者建立起一座沟通数学原理和互联网实际业务应用的桥梁。同时本书给出了案例中的完整代码以及分析过程，力图帮助读者充分理解R语言是如何实现算法的。

《R语言与网站分析》是目前为止国内唯一一本阐述如何使用R语言来分析和挖掘互联网数据的应用性书籍。书中使用大量的实际案例，把数学原理同R语言实现方案有机结合起来。力图通过案例分析达到举一反三的效果，进而指导读者在日后的实际工作中进行应用。同时书中也阐述众多常用的数据分析和挖掘的方法和原理，对于非互联网的从业人员也很有指导意义

# 《R语言与网站分析》

## 作者简介

李明，曾就读于锦州市锦州中学 毕业于沈阳理工大学 信息工程学院 电子科技与技术系，曾就职于凡客诚品、居然之家等大型电子商务公司。目前就职于优酷土豆网。研究方向是R语言同互联网数据分析/挖掘的结合。撰写过大量R语言的基础和高级应用类文章，对互联网数据统计系统的R语言实践有较深研究。

联系方式：博客[www.bassary.com](http://www.bassary.com)；微博 [www.weibo.com/2382348263](http://www.weibo.com/2382348263)。

## 书籍目录

### 前言

### 第1章 统计思维与网站分析 1

#### 1.1 不确定与确定 1

#### 1.2 统计分析方法 4

##### 1.2.1 细分分析 4

##### 1.2.2 对比分析 6

##### 1.2.3 趋势分析 7

#### 1.3 网站分析概要 7

##### 1.3.1 解决用户需求 7

##### 1.3.2 寻找新的用户需求 10

### 第2章 R语言数据操作基础 11

#### 2.1 R简介 11

#### 2.2 了解R软件 11

##### 2.2.1 软件安装 11

##### 2.2.2 R软件界面 12

##### 2.2.3 工作目录 14

##### 2.2.4 命令行交互 15

##### 2.2.5 命令脚本文件 16

##### 2.2.6 工作空间数据 16

##### 2.2.7 帮助 17

##### 2.2.8 R语言入门 18

##### 2.2.9 扩展算法包 19

#### 2.3 R语言的数据类型 21

#### 2.4 对象及其属性 23

##### 2.4.1 固有属性：模式和长度 24

##### 2.4.2 读取和设置属性值 24

##### 2.4.3 对象的搜索和删除 25

#### 2.5 向量 26

##### 2.5.1 创建向量 26

##### 2.5.2 向量索引 26

##### 2.5.3 向量编辑 29

##### 2.5.4 向量排序 29

##### 2.5.5 向量去重 30

##### 2.5.6 缺失值处理 30

##### 2.5.7 向量间操作 31

#### 2.6 矩阵和数组 31

##### 2.6.1 创建矩阵 32

##### 2.6.2 矩阵索引 33

##### 2.6.3 矩阵编辑 34

##### 2.6.4 矩阵的运算 35

##### 2.6.5 数组 37

##### 2.6.6 apply函数应用 38

#### 2.7 列表和数据框 39

##### 2.7.1 列表的创建和索引 39

##### 2.7.2 列表编辑 40

##### 2.7.3 数据框的创建和名称 40

##### 2.7.4 数据框索引 41

- 2.7.5 数据框编辑 43
- 2.7.6 缺失值处理 44
- 2.8 因子 44
  - 2.8.1 无序和有序因子 45
  - 2.8.2 连续数据的离散化 46
- 2.9 字符串操作 47
  - 2.9.1 字符串长度nchar 48
  - 2.9.2 字符串合并和分割 48
  - 2.9.3 字符串内部字符的读取和替换 49
  - 2.9.4 正则表达式 52
- 2.10 常用数据的创建 53
  - 2.10.1 因子序列的创建 53
  - 2.10.2 等差序列的创建 54
  - 2.10.3 随机抽样sample 54
  - 2.10.4 重复序列rep 55
  - 2.10.5 概率分布 55
- 2.11 控制流 58
  - 2.11.1 分支语句 58
  - 2.11.2 循环语句 59
- 2.12 运算符、函数和过程 61
  - 2.12.1 常用运算符 61
  - 2.12.2 函数 62
  - 2.12.3 过程 63
- 2.13 数据的读写等操作 64
  - 2.13.1 读取数据 64
  - 2.13.2 输出数据 68
- 第3章 R语言的绘图基础 69
  - 3.1 概述 69
  - 3.2 颜色以及文字/点/线参数的设置 69
    - 3.2.1 颜色 71
    - 3.2.2 文字元素相关参数设置 76
    - 3.2.3 点元素相关参数设置 77
    - 3.2.4 线元素相关参数设置 79
  - 3.3 低级绘图函数 80
    - 3.3.1 引例 80
    - 3.3.2 标题 81
    - 3.3.3 图例 83
    - 3.3.4 坐标轴 84
    - 3.3.5 边框 88
    - 3.3.6 网格线 89
    - 3.3.7 点 90
    - 3.3.8 线 91
    - 3.3.9 文字 94
    - 3.3.10 多边形 96
  - 3.4 高级绘图函数以及常用绘图应用 96
    - 3.4.1 图形类型的选择 96
    - 3.4.2 散点图 98
    - 3.4.3 气泡图 100
    - 3.4.4 曲线图 102

- 3.4.5 柱状图 105
- 3.4.6 条形图 109
- 3.4.7 饼图 111
- 3.4.8 面积堆积图 115
- 3.4.9 直方图和密度曲线图 118
- 3.5 绘图窗口操作函数 120
  - 3.5.1 单一窗口中的子绘图区域布局 120
  - 3.5.2 绘图窗口操作 120
- 第4章 单指标分析 121
  - 4.1 指标描述 121
    - 4.1.1 平均值和集中趋势 121
    - 4.1.2 正态分布 126
    - 4.1.3 频数分析 130
    - 4.1.4 描述性分析指标 131
  - 4.2 异常点监控 133
    - 4.2.1 概述 133
    - 4.2.2 P控制图：监控转化率型指标 135
    - 4.2.3 单值-均值控制图 142
    - 4.2.4 单值-移动极差控制图 147
  - 4.3 连续型指标的对比 150
    - 4.3.1 数据变换 150
    - 4.3.2 假设检验 152
    - 4.3.3 相同指标内的两组数对比：T检验 153
    - 4.3.4 相同指标内的多组数据对比：单因素方差分析 156
    - 4.3.5 单因素协方差分析 163
  - 4.4 分类型指标的对比 167
    - 4.4.1 列联表分析 167
    - 4.4.2 卡方独立性检验 172
- 第5章 时间序列分析 177
  - 5.1 时间序列 177
  - 5.2 增长率 180
    - 5.2.1 环比增长率 180
    - 5.2.2 同比增长率 181
  - 5.3 移动平均 181
    - 5.3.1 数学原理概述 181
    - 5.3.2 filter函数 182
    - 5.3.3 R语言实现 184
  - 5.4 指数平滑 185
    - 5.4.1 一次指数平滑 186
    - 5.4.2 二次指数平滑 187
    - 5.4.3 三次指数平滑 191
  - 5.5 ARIMA模型 194
    - 5.5.1 自相关性 194
    - 5.5.2 平稳性和白噪声 199
    - 5.5.3 MA滑动平均过程 200
    - 5.5.4 AR自回归过程 202
    - 5.5.5 ARMA自回归滑动平均混合过程 203
    - 5.5.6 检验模型质量 205
    - 5.5.7 非平稳时间序列的ARIMA过程 208

- 第6章 连续指标建模：回归分析 213
  - 6.1 一元线性回归分析 213
    - 6.1.1 引例 213
    - 6.1.2 一元线性回归分析的原理及R语言实现 214
  - 6.2 多元回归分析 221
    - 6.2.1 引例 222
    - 6.2.2 多元线性回归分析建模 222
    - 6.2.3 模型修正函数update() 223
    - 6.2.4 逐步回归分析函数step() 226
    - 6.2.5 自变量中包含分类型数据的回归分析 228
  - 6.3 Logic回归分析 230
    - 6.3.1 引例及数据 230
    - 6.3.2 logic分析的原理 230
    - 6.3.3 R语言实现 232
  - 6.4 回归树CART 237
    - 6.4.1 rpart函数 237
    - 6.4.2 预测及模型性能衡量 240
    - 6.4.3 过度拟合和剪枝 240
- 第7章 分类指标建模：分类分析 243
  - 7.1 决策树分类分析 243
    - 7.1.1 概述 243
    - 7.1.2 C4.5算法 247
    - 7.1.3 CART算法 258
    - 7.1.4 条件推理决策树算法 265
    - 7.1.5 随机森林算法 267
  - 7.2 贝叶斯分类 268
    - 7.2.1 贝叶斯定理 268
    - 7.2.2 朴素贝叶斯分类器 270
  - 7.3 支持向量机SVM 271
    - 7.3.1 原理 271
    - 7.3.2 在R语言中实现非线性SVM分析 273
  - 7.4 人工神经网络 274
    - 7.4.1 神经元 274
    - 7.4.2 两层网络 275
    - 7.4.3 反向传播算法 276
    - 7.4.4 R语言实现 282
    - 7.4.5 隐藏层中神经单元数目的确定 283
  - 7.5 分类器的性能评估 284
    - 7.5.1 混淆矩阵 284
    - 7.5.2 ROC曲线和AUC 287
    - 7.5.3 提升度和提升曲线 291
    - 7.5.4 洛伦兹曲线 292
- 第8章 样本细分 293
  - 8.1 数据降维 293
    - 8.1.1 问题引入 293
    - 8.1.2 因子分析概述 294
    - 8.1.3 factanal函数 296
    - 8.1.4 实例：问卷调查的因子分析 297
  - 8.2 聚类分析 303

- 8.2.1 距离 303
- 8.2.2 层次聚类 313
- 8.2.3 kmeans聚类 318
- 8.3 样本判别 321
  - 8.3.1 knn (k最近邻分类) 算法 323
  - 8.3.2 实例：基于knn算法的商品推荐系统 324
- 第9章 样本间的关系 329
  - 9.1 关联分析 329
    - 9.1.1 关联分析的原理 329
    - 9.1.2 Apriori算法的R语言实现 331
  - 9.2 序列模式关联分析 334
    - 9.2.1 序列模式关联分析的原理 334
    - 9.2.2 序列模式关联分析的R语言实现 336
    - 9.2.3 实例：用户点击页面的行为分析 338
  - 9.3 关系网络分析 346
    - 9.3.1 网络关系分析的基本概念 346
    - 9.3.2 中心度 356
    - 9.3.3 中心势 362
    - 9.3.4 社群发现 363
    - 9.3.5 实例：分析微博的传播特性 367
    - 9.3.6 实例：购物车商品分类分析 370
- 第10章 文本分析 378
  - 10.1 数据处理 378
    - 10.1.1 数据引例 378
    - 10.1.2 分词和词条 380
    - 10.1.3 语料库和文档 381
    - 10.1.4 词条—文档关系矩阵 384
  - 10.2 实例：QQ群消息历史分析 387
    - 10.2.1 数据集 387
    - 10.2.2 数据集格式转化 389
    - 10.2.3 留言文本的分词操作 390
    - 10.2.4 分析常用话题词汇 391
    - 10.2.5 建立用户—词条间的网络关系图 392
    - 10.2.6 绘制重点词条和用户的网络图 392
  - 10.3 XML包爬取网页数据 394
    - 10.3.1 htmlParse函数 395
    - 10.3.2 getNodeSet函数 395
    - 10.3.3 xmlValue函数 396
    - 10.3.4 xmlGetAttr函数 396
    - 10.3.5 实例：爬取某电商网站衬衫类目商品的部分信息 397
- 第11章 网站指标监控系统的搭建 400
  - 11.1 gWidgets包基础 400
    - 11.1.1 环境搭建 400
    - 11.1.2 引例 401
    - 11.1.3 常用控件 406
  - 11.2 实例：商品搜索算法参数配置监控GUI系统 414
    - 11.2.1 商品搜索算法概述 414
    - 11.2.2 原始商品信息 415
    - 11.2.3 系统概述 417



- 11.2.4 模块布局 418
- 11.2.5 计算除关键字匹配以外的指标得分 420
- 11.2.6 模拟一次搜索行为数据 421
- 11.2.7 模拟多次搜索行为数据 423
- 11.2.8 绘图 423
- 第12章 基于RFM模型的客户价值系统 427
  - 12.1 马尔科夫链 427
    - 12.1.1 引例 427
    - 12.1.2 概念 428
    - 12.1.3 预测实例 428
  - 12.2 RFM模型 431
    - 12.2.1 RFM的概念 431
    - 12.2.2 数据集 431
    - 12.2.3 购买行为随机模型 432
  - 12.3 结合马尔科夫链的RFM模型 436
    - 12.3.1 购买状态转移矩阵 437
    - 12.3.2 预测用户购买状态 445

# 《R语言与网站分析》

## 精彩短评

- 1、书质量很差，套着网站分析外壳的R基础教程，而且基础讲的完全不如R的基础教程书。书中的案例感觉都是模拟的数据，丝毫看不出实用性，因此分析方法和分析角度最多也就是个理论上正确，看着很没意思。
- 2、只看了第2章R语言基础，按对象类型讲操作，比较清晰。
- 3、写的还是挺实在的,支持原创R图书
- 4、很不错哟~
- 5、读的特别快，觉得还是蛮实用的。
- 6、还行，用了点心思写
- 7、看了下样章，且不说将qcc错写成gcc这样的错误。代码也是虎头蛇尾，看着让人窝心！！！！

# 《R语言与网站分析》

## 精彩书评

1、不知道是自己长进太快了，还是这书太傻B了，反正翻书的节奏加快了，都有些想放弃那好几册子machine learning了

# 《R语言与网站分析》

## 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)