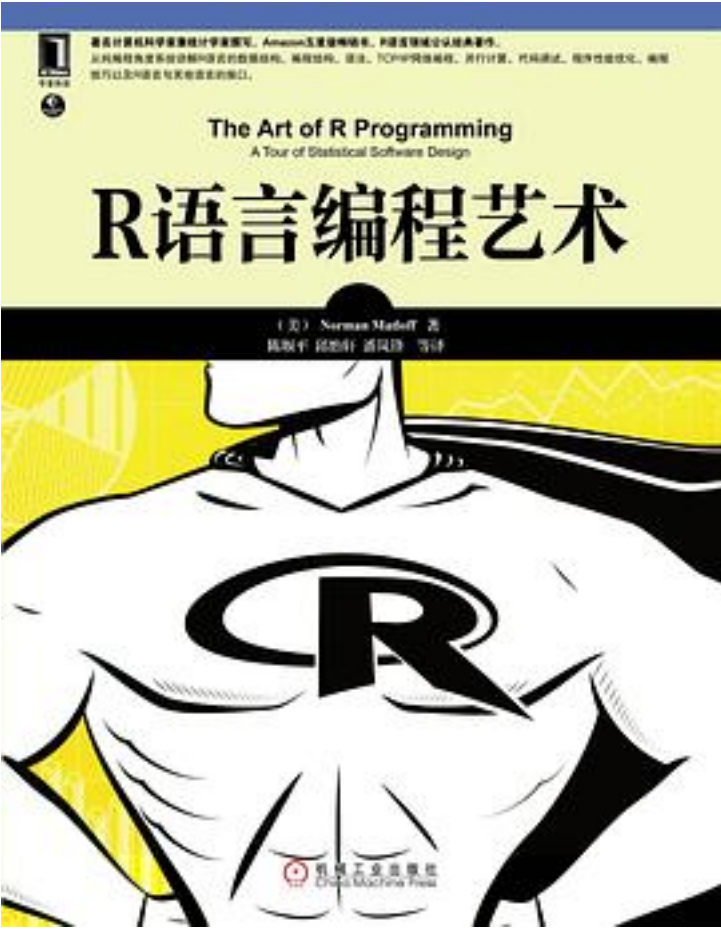


R语言编程艺术



[R语言编程艺术_下载链接1](#)

著者: (美) Norman Matloff

出版者:机械工业出版社

出版时间:2013-5

装帧:平装

isbn:9787111423140

【编辑推荐】

这本书涵盖了R语言编程的诸多方面，尤其在面向对象编程、程序调试、提升程序运行

速度以及并行计算等方面，填补了同类图书的空白。关于程序调试的章节更是作者多年经验的总结。不管是初学者还是有一定编程经验的读者，阅读这本书都会有所收获。

——统计之都

【内容简介】

R语言是世界上最流行的用于数据处理和统计分析的脚本语言。考古学家用它来跟踪古代文明的传播，医药公司用它来探索哪种药物更安全、更有效，精算师用它评估金融风险以保证市场的平稳运行。总之，在大数据时代，统计数据、分析数据都离不开计算机软件的支持，在这方面R语言尤其出色。

本书将带领你踏上R语言软件开发之旅，从最基本的数据类型和数据结构开始，到闭包、递归和匿名函数等高级主题，由浅入深，讲解细腻，读者完全不需要统计学的知识，甚至不需要编程基础。而书中提到的很多高级编程技巧，都是作者多年编程经验的总结，对有经验的开发者也大有裨益。本书精选了44个扩展案例，这些案例都源自于作者亲身参与过的咨询项目，都是与数据分析相关的，生动展示了R语言在统计学中的高效应用。

本书核心内容：

R语言的完整语法以及R语言的编程技巧

创建精美图形来展示复杂数据和函数

使用并行计算和向量化的方法编写更高效的代码

使用R对C/C++和Python的接口来提高计算速度或增加功能

文本分析、图像处理等领域新的R包

使用高级调试技巧清除代码里恼人的错误

包含许多“扩展案例”，展示完整的、特定用途的函数，并针对同一个问题讨论了不同的设计方案，以便分析高效准确的做法。

在恰当的时候介绍R语言与其他语言的差异，给那些了解其他语言的开发人员提供参考。

【相关图书推荐】

978-7-111-40700-3 数据挖掘与R语言

978-7-111-42021-7 R语言经典实例

978-7-111-41731-6 机器学习：实用案例解析

978-7-111-32572-7 时间序列分析及应用：R语言（原书第2版）

作者介绍:

Norman Matloff

著名计算机科学家兼统计学家，美国加州大学戴维斯分校计算机科学系教授，曾是该校统计专业的创建者之一，并担任过统计学教授。对并行编程、网络流量、数据挖掘、磁盘系统性能等方面的技术都有深入的研究。乐于分享，撰写了多部广受欢迎的关于软件开发的在线教程，多次为《纽约时报》、《华盛顿邮报》、《福布斯杂志》以及《洛杉矶时报》撰写文章，同时他还是《The Art of Debugging》的作者之一。

目录: 译者序

前言

致谢

第1章 快速入门

1.1 怎样运行R

1.1.1 交互模式

1.1.2 批处理模式

1.2 第一个R会话

1.3 函数入门

1.3.1 变量的作用域

1.3.2 默认参数

1.4 R语言中一些重要的数据结构

1.4.1 向量，R语言中的战斗机

1.4.2 字符串

1.4.3 矩阵

1.4.4 列表

1.4.5 数据框

1.4.6 类

1.5 扩展案例：考试成绩的回归分析

1.6 启动和关闭R

1.7 获取帮助

1.7.1 help()函数

1.7.2 example()函数

1.7.3 如果你不太清楚要查找什么

1.7.4 其他主题的帮助

1.7.5 批处理模式的帮助

1.7.6 互联网资源

第2章 向量

2.1 标量、向量、数组与矩阵

2.1.1 添加或删除向量元素

2.1.2 获取向量长度

2.1.3 作为向量的矩阵和数组

2.2 声明

2.3 循环补齐

2.4 常用的向量运算

2.4.1 向量运算和逻辑运算

2.4.2 向量索引

2.4.3 用：运算符创建向量

2.4.4 使用seq()创建向量

2.4.5 使用rep()重复向量常数

2.5 使用all()和any()

2.5.1 扩展案例：寻找连续出现1的游程

2.5.2 扩展案例：预测离散值时间序列

2.6 向量化运算符

2.6.1 向量输入，向量输出

2.6.2 向量输入，矩阵输出

- 2.7 NA与NULL值
 - 2.7.1 NA的使用
 - 2.7.2 NULL的使用
- 2.8 筛选
 - 2.8.1 生成筛选索引
 - 2.8.2 使用subset()函数筛选
 - 2.8.3 选择函数which()
- 2.9 向量化的ifelse()函数
 - 2.9.1 扩展案例：度量相关性
 - 2.9.2 扩展案例：对鲍鱼数据集重新编码
- 2.10 测试向量相等
- 2.11 向量元素的名称
- 2.12 关于c()的更多内容
- 第3章 矩阵和数组
 - 3.1 创建矩阵
 - 3.2 一般矩阵运算
 - 3.2.1 线性代数运算
 - 3.2.2 矩阵索引
 - 3.2.3 扩展案例：图像操作
 - 3.2.4 矩阵元素筛选
 - 3.2.5 扩展案例：生成协方差矩阵
- 3.3 对矩阵的行和列调用函数
 - 3.3.1 使用apply()函数
 - 3.3.2 扩展案例：寻找异常值
- 3.4 增加或删除矩阵的行或列
 - 3.4.1 改变矩阵的大小
 - 3.4.2 扩展案例：找到图中距离最近的一对端点
- 3.5 向量与矩阵的差异
- 3.6 避免意外降维
- 3.7 矩阵的行和列的命名问题
- 3.8 高维数组

- 第4章 列表
 - 4.1 创建列表
 - 4.2 列表的常规操作
 - 4.2.1 列表索引
 - 4.2.2 增加或删除列表元素
 - 4.2.3 获取列表长度
 - 4.2.4 扩展案例：文本词汇索引
 - 4.3 访问列表元素和值
 - 4.4 在列表上使用apply系列函数
 - 4.4.1 lapply()和sapply()的使用
 - 4.4.2 扩展案例：文本词汇索引（续）
 - 4.4.3 扩展案例：鲍鱼数据
 - 4.5 递归型列表
- 第5章 数据框
 - 5.1 创建数据框
 - 5.1.1 访问数据框
 - 5.1.2 扩展案例：考试成绩的回归分析（续）
 - 5.2 其他矩阵式操作
 - 5.2.1 提取子数据框
 - 5.2.2 缺失值的处理
 - 5.2.3 使用rbind()和cbind()等函数
 - 5.2.4 使用apply()
 - 5.2.5 扩展案例：工资研究

- 5.3 合并数据框
- 5.4 应用于数据框的函数
 - 5.4.1 在数据框上应用lapply()和sapply()函数
 - 5.4.2 扩展案例：应用Logistic模型
 - 5.4.3 扩展案例：学习中文方言的辅助工具
- 第6章 因子和表
 - 6.1 因子与水平
 - 6.2 因子的常用函数
 - 6.2.1 tapply函数
 - 6.2.2 split()函数
 - 6.2.3 by()函数
 - 6.3 表的操作
 - 6.3.1 表中有关矩阵和类似数组的操作
 - 6.3.2 扩展案例：提取子表
 - 6.3.3 扩展案例：在表中寻找频数最大的单元格
 - 6.4 其他与因子和表有关的函数
 - 6.4.1 aggregate()函数
 - 6.4.2 cut()函数
- 第7章 R语言编程结构
 - 7.1 控制语句
 - 7.1.1 循环
 - 7.1.2 对非向量集合的循环
 - 7.1.3 if-else结构
 - 7.2 算术和逻辑运算符及数值
 - 7.3 参数的默认值
 - 7.4 返回值
 - 7.4.1 决定是否显式调用return ()
 - 7.4.2 返回复杂对象
 - 7.5 函数都是对象
 - 7.6 环境和变量作用域的问题
 - 7.6.1 顶层环境
 - 7.6.2 变量作用域的层次
 - 7.6.3 关于ls()的进一步讨论
 - 7.6.4 函数（几乎）没有副作用
 - 7.6.5 扩展案例：显示调用框的函数
 - 7.7 R语言中没有指针
 - 7.8 向上级层次进行写操作
 - 7.8.1 利用超赋值运算符对非局部变量进行写操作
 - 7.8.2 用assign()函数对非局部变量进行写操作
 - 7.8.3 扩展案例：用R语言实现离散事件仿真
 - 7.8.4 什么时候使用全局变量
 - 7.8.5 闭包
 - 7.9 递归
 - 7.9.1 Quicksort的具体实现
 - 7.9.2 拓展举例：二叉查找树
 - 7.10 置换函数
 - 7.10.1 什么是置换函数
 - 7.10.2 扩展案例：可记录元素修改次数的向量类
 - 7.11 写函数代码的工具
 - 7.11.1 文本编辑器和集成开发环境
 - 7.11.2 edit()函数
 - 7.12 创建自己的二元运算符
 - 7.13 匿名函数
- 第8章 数学运算与模拟

- 8.1 数学函数
 - 8.1.1 扩展例子：计算概率
 - 8.1.2 累积和与累积乘积
 - 8.1.3 最小值和最大值
 - 8.1.4 微积分
- 8.2 统计分布函数
- 8.3 排序
- 8.4 向量和矩阵的线性代数运算
 - 8.4.1 扩展示例：向量叉积
 - 8.4.2 扩展示例：确定马尔科夫链的平稳分布
- 8.5 集合运算
- 8.6 用R做模拟
 - 8.6.1 内置的随机变量发生器
 - 8.6.2 重复运行时获得相同的随机数流
 - 8.6.3 扩展案例：组合的模拟
- 第9章 面向对象的编程
 - 9.1 S3类
 - 9.1.1 S3泛型函数
 - 9.1.2 实例：线性模型函数lm()中的OOP
 - 9.1.3 寻找泛型函数的实现方法
 - 9.1.4 编写S3类
 - 9.1.5 使用继承
 - 9.1.6 扩展示例：用于存储上三角矩阵的类
 - 9.1.7 扩展示例：多项式回归程序
 - 9.2 S4类
 - 9.2.1 编写S4类
 - 9.2.2 在S4类上实现泛型函数
 - 9.3 S3类和S4类的对比
 - 9.4 对象的管理
 - 9.4.1 用ls()函数列出所有对象
 - 9.4.2 用rm()函数删除特定对象
 - 9.4.3 用save()函数保存对象集合
 - 9.4.4 查看对象内部结构
 - 9.4.5 exists()函数
- 第10章 输入与输出
 - 10.1 连接键盘与显示器
 - 10.1.1 使用scan()函数
 - 10.1.2 使用readline()函数
 - 10.1.3 输出到显示器
 - 10.2 读写文件
 - 10.2.1 从文件中读取数据框或矩阵
 - 10.2.2 读取文本文件
 - 10.2.3 连接的介绍
 - 10.2.4 扩展案例：读取PUMS普查数据
 - 10.2.5 通过URL在远程计算机上访问文件
 - 10.2.6 写文件
 - 10.2.7 获取文件和目录信息
 - 10.2.8 扩展案例：多个文件内容的和
 - 10.3 访问互联网
 - 10.3.1 TCP/IP概述
 - 10.3.2 R中的socket
 - 10.3.3 扩展案例：实现R的并行计算
- 第11章 字符串操作
 - 11.1 字符串操作函数概述

- 11.1.1 grep()
- 11.1.2 nchar()
- 11.1.3 paste()
- 11.1.4 sprintf()
- 11.1.5 substr()
- 11.1.6 strsplit()
- 11.1.7 regexpr()
- 11.1.8 gregexpr()
- 11.2 正则表达式
 - 11.2.1 扩展案例：检测文件名的后缀
 - 11.2.2 扩展案例：生成文件名
- 11.3 在调试工具edtdbg中使用字符串工具

第12章 绘图

- 12.1 创建图形
 - 12.1.1 基础图形系统的核心：plot()函数
 - 12.1.2 添加线条：abline()函数
 - 12.1.3 在保持现有图形的基础上新增一个绘图窗口
 - 12.1.4 扩展案例：在一张图中绘制两条密度曲线
 - 12.1.5 扩展案例：进一步考察多项式回归
 - 12.1.6 添加点：points()函数
 - 12.1.7 添加图例：legend()函数
 - 12.1.8 添加文字：text()函数
 - 12.1.9 精确定位：locator()函数
 - 12.1.10 保存图形
- 12.2 定制图形
 - 12.2.1 改变字符大小：cex选项
 - 12.2.2 改变坐标轴的范围：xlim和ylim选项
 - 12.2.3 添加多边形：polygon()函数
 - 12.2.4 平滑散点：lowess()和loess()函数
 - 12.2.5 绘制具有显式表达式的函数
 - 12.2.6 扩展案例：放大曲线的一部分
- 12.3 将图形保存到文件
 - 12.3.1 R图形设备
 - 12.3.2 保存已显示的图形
 - 12.3.3 关闭R图形设备
- 12.4 创建三维图形

第13章 调试

- 13.1 调试的基本原则
 - 13.1.1 调试的本质：确认原则
 - 13.1.2 从小处着手
 - 13.1.3 模块化的、自顶向下的调试风格
 - 13.1.4 反漏洞
- 13.2 为什么要使用调试工具
- 13.3 使用R的调试工具
 - 13.3.1 利用debug()和browser()函数进行逐步调试
 - 13.3.2 使用浏览器命令
 - 13.3.3 设置断点
 - 13.3.4 使用trace()函数进行追踪
 - 13.3.5 使用traceback()和debugger()函数对崩溃的程序进行检查
 - 13.3.6 扩展案例：两个完整的调试会话
- 13.4 更方便的调试工具
- 13.5 在调试模拟数据的代码时请确保一致性
- 13.6 语法和运行时错误
- 13.7 在R上运行GDB

第14章 性能提升：速度和内存

14.1 编写快速的R代码

14.2 可怕的for循环

14.2.1 用向量化提升速度

14.2.2 扩展案例：在蒙特卡罗模拟中获得更快的速度

14.2.3 扩展案例：生成幂次矩阵

14.3 函数式编程和内存问题

14.3.1 向量赋值问题

14.3.2 改变时拷贝

14.3.3 扩展案例：避免内存拷贝

14.4 利用Rprof()来寻找代码的瓶颈

14.4.1 利用Rprof()来进行监视

14.4.2 Rprof()的工作原理

14.5 字节码编译

14.6 内存无法装下数据怎么办

14.6.1 分块

14.6.2 利用R软件包来进行内存管理

第15章 R与其他语言的接口

15.1 编写能被R调用的C/C++函数

15.1.1 R与C/C++交互的预备知识

15.1.2 例子：提取方阵的次对角线元素

15.1.3 编译和运行程序

15.1.4 调试R/C程序

15.1.5 扩展案例：预测离散取值的时间序列

15.2 从Python调用R

15.2.1 安装RPy

15.2.2 RPy语法

第16章 R语言并行计算

16.1 共同外链问题

16.2 snow包简介

16.2.1 运行snow代码

16.2.2 分析snow代码

16.2.3 可以获得多少倍的加速

16.2.4 扩展案例：K均值聚类

16.3 借助于C

16.3.1 利用多核机器

16.3.2 扩展案例：利用OpenMP解决共同外链问题

16.3.3 运行OpenMP代码

16.3.4 OpenMP代码分析

16.3.5 其他OpenMP指令

16.3.6 GPU编程

16.4 普遍的性能考虑

16.4.1 开销的来源

16.4.2 简单并行程序，以及那些不简单的

16.4.3 静态和动态任务分配

16.4.4 软件炼金术：将一般的问题转化为简单并行问题

16.5 调试R语言并行计算的代码

附录A 安装R

附录B 安装和使用包

• • • • • ([收起](#))

标签

R

数据分析

R语言

编程

数据挖掘

计算机

程序设计

统计

评论

R的迭代速度太快了。这本书的内容已经大部分过时了并且它包含了太多不需要的细节。对于习惯传统的编码形式并且想要转移到R上的程序员，这本书仍然是一本不可多得的佳作。而对于数据科学家来说，这本书只需要快速翻过就好。

查漏补缺来的。比较适合想了解R的计算机专业人士入门学习，书中更多是介绍语言本身，涉及统计知识比较少；如果是统计学专业，对编程语言了解比较少的，更建议先读R语言实战，然后再浏览这本书，或者混着看吧。总之是基础知识了。|
笔记做完了，080201。

翻译的很好

写给程序员的R语言书

: TP312/5423

用到了一点点。。

数据结构讲解清楚应该是最大的优点，最后几章超棒——少有的中文R相关资料——第二大优点。中译版提供的勘误和code弥补了英文版的相关不足：<https://github.com/consname/art-r-translation> 作为一本语言书，已经很赞了。

最赞的是case study部分的内容。可以定位为一本“面向程序员而不是分析师的R语言教材”，但是感觉很多东西没有涉及到，难道只能去翻阅文档了么？

这本书是按照编程的目的编写的，从数据结构，到控制语句，再到面向对象、调试和优化。。。通读了一遍，一开始有点难度，收获很大，还需要多多实践案例，它也可以当工具书来用的~~

我的R入门书。非常不错。就是讲绘图少了点。

没读完，太贵了~

入门来说优于 r in action

很好的入门书

需要做一个 talk，又翻了一遍中文版，非常适合码农读

很不错，简单清楚，适合入门

上次装逼的时候看过

我还是得找个东西亲手做做才上手得快啊

这本书从语言本身介绍了R的一系列特征，详细介绍了R中的常见数据类型，面向对象，字符串处理，绘图，调试与性能，并行计算等。书中知识点都很纯粹，没有涉及到数据科学中复杂的数学模型和相关算法。如果你想要全面了解R语言，建议读这本书。

结构非常清晰，从语言处理的各类元素展开解释，再介绍基础的操作。难度适中的入门读物，能对R的构架有比较扎实的打底。

先拜下大神 暑假必然好好钻研..

[R语言编程艺术 下载链接1](#)

书评

写于 <http://datawarehou.se/comments/the-art-of-r-programming/>
基本看完了这本《R语言艺术》。该书并没有中文版，我在Amazon买的Kindle版来看。
最初是在flowing

data上面看的书评，留下了印象，而本身自己也挺想了解下R，索性就通过这本书来入门，应该说目的还是达到了。后...

这本书适用于各阶段的读者，尤其是会一点R然后搞不清楚 data structures之间区别的读者（其实就是我！）没系统学习data structures之前总是出现莫名其妙的bug，其实只需要给factor加上as.numeric就能转化成数字了，然后进行数字运算。讲解易懂，简洁，是本好书。

花了大致两周半的时间，断断续续地读完了这本「R语言的艺术」正如这本书的副标题所说，书的内容主要是 A Tour of Statistical Software Design。所以，如果你是一个软件工程师，对于R语言感兴趣，对于统计学的原理或许不是特别熟悉。那么，通过阅读这本书，你可以得到R语言在...

这是我所读的R系列中的最佳读本！

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...

不确定作者有没有在认真对待这本书的写作。我身为一个统计专业，同时搞算法研究的学生，在编程和统计两方面应该都没有瘸腿，但这本书根本没有让我对R的了解有更多的深入。
书对简单易懂的东西谈得太多，而对那些生僻的东西谈得太少。生僻的东西很多都是读者感兴趣的，像是R与C...

[R语言编程艺术_下载链接1](#)