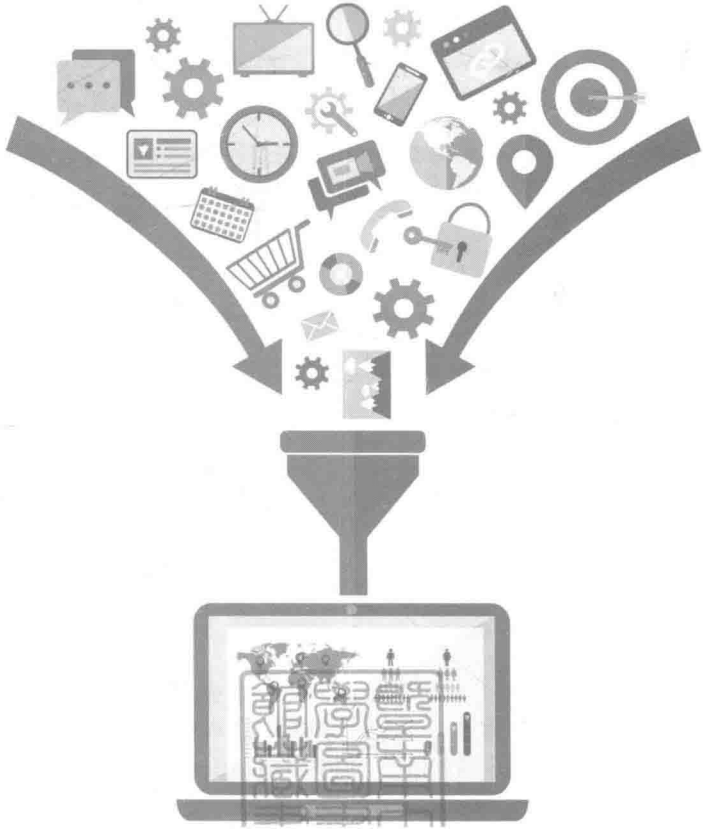
Julia
数据科学应用

[美] 扎卡赖亚斯·弗格里斯（Zacharias Voulgaris） 著 陈光欣 译

中国工信出版集团
人民邮电出版社
POSTS & TELECOM PRESS

Julia
数据科学应用

[美]扎卡赖亚斯·弗格里斯（Zacharias Voulgaris） 著 陈光欣 译



人 民 邮 电 出 版 社
北 京

图书在版编目（CIP）数据

Julia数据科学应用 /（美）弗格里斯（Zacharias Voulgaris）著；陈光欣译. -- 北京：人民邮电出版社，2018.2
ISBN 978-7-115-47328-8

Ⅰ. ①J… Ⅱ. ①弗… ②陈… Ⅲ. ①程序语言－研究 Ⅳ. ①TP312

中国版本图书馆CIP数据核字(2017)第309720号

版权声明

◆ 著　[美] 扎卡赖亚斯·弗格里斯（Zacharias Voulgaris）
译　陈光欣
责任编辑　陈冀康
责任印制　焦志炜

◆ 人民邮电出版社出版发行　北京市丰台区成寿寺路 11 号
邮编　100164　电子邮件　315@ptpress.com.cn
网址　http://www.ptpress.com.cn
大厂聚鑫印刷有限责任公司印刷

◆ 开本：720×960　1/16
印张：19.25
字数：240 千字　2018 年 2 月第 1 版
印数：1－2 400 册　2018 年 2 月河北第 1 次印刷

著作权合同登记号　图字：01-2016-8081 号

定价：69.00 元

读者服务热线：(010) 81055410　印装质量热线：(010) 81055316
反盗版热线：(010) 81055315
广告经营许可证：京东工商广登字 20170147 号

内容提要

数据科学通过各种统计学和机器学习的技术与方法，将数据转换为有用的信息或知识。Julia 是一种在数据科学领域逐渐流行起来的语言。

本书会提出一系列在数据科学流程中常见的、有代表性的实际问题，并指导读者使用 Julia 去解决这些问题。全书共 13 章，涵盖了 Julia 基础知识、工作环境搭建、语言基础和高级内容、数据科学应用、数据可视化、机器学习方法（包括无监督式学习方法和监督式学习方法）、图分析方法等重要的话题。附录部分给出了学习和使用 Julia 的一些有用的参考资料，还给出了各章的思考题的答案。

本书适合对数据科学的知识和应用方法感兴趣的读者阅读，特别适合有志于学习 Julia 并从事数据科学相关工作的人员学习参考。

前言

我是在几年前发现 Julia 的，从此就被它的强大能力与巨大潜力所吸引。Julia 具有用户友好的集成开发环境（Integrated Development Environment，IDE），这使它很容易上手；它还具有高级的逻辑表达能力（非常类似 Matlab 和其他高级语言）和极高的性能，这使它的功能非常强大。但是，当时我正致力于研究其他更成熟的平台，比如 R 和 Java，未能给予 Julia 太多的关注。

因此，我只了解了 Julia 的一些基础知识，以及当时的教程中提供的一些具体应用，并没有进行更多的研究。除了 Julia 之外，我还知道不断有一些新的有趣的语言被开发出来，但大多数是昙花一现。

那么，为什么我现在又对 Julia 感兴趣了呢？一个原因就是，这些年它一直保持着良好的发展势头，Julia 会议的参与人数每年都有显著的增长。尽管我曾经很熟悉它的基本知识，但当我重拾 Julia 时，发现有很多新的知识需要学习。从我初识 Julia 之后，它已经有了很大的发展。

更重要的原因是，Julia 已经跨过了大西洋，引起了欧洲从业者的极大兴趣，其中一位已经为这种相当年轻的语言创建了一系列视频和练习资料。

在试用了 Julia 0.2 版之后，我开始琢磨，除了快速分解质因数和计算第 n 个斐波那契数之外，是否能使用 Julia 来做些真正有用的事情。虽然 0.2 版仅有几个软件包，文档也做得很差，我只能找到零星几个介绍这门语言的视频，多数还是来自某个 Python 会议上的发言。但是，我还是在计算机上保留着 Julia，并时不时地用它写个脚本，来解决 Project Euler、Programming Praxis 或类似站点上的编程问题。当时我是个项目经理，所以没有很大的积极性去掌握一门新的编程语言。我在 Julia 上所做的一切都是出于兴趣。

但是，几个月之后，我重新开始从事数据科学工作，并更加正式地使用 Julia 编程。我很快就发现，使用 Julia 编写代码比使用 Python 更容易，例如，即使使

用 Python 完成一个基本的数据加工任务，也需要一大堆扩展包。

在使用 Julia 解决小问题之后，我决定使用 Julia 独立完成一个完整的数据科学项目。在经历了不可避免的学习曲线和成长阵痛之后，我终于达到了目标。这并不是我最得意的成果，但它证明了在进行一些训练、尝试和纠错之后，Julia 可以高效地完成正式的数据科学任务。

在本书中，我会分享在这个项目以及随后的项目中获得的经验，阐述如何在数据科学的各个环节使用 Julia。尽管现在已经有了一些介绍 Julia 的书籍，但还没有一本全面介绍如何在数据科学领域内应用 Julia 的专著。我曾非常期待有这样一本书，但有了多年使用 Julia 的经验之后，我决定亲自上阵，撰写这样的一本书。

我完全清楚，撰写一本介绍正处于发展时期的语言的书风险有多大，但是 Julia 这门语言不会停止发展，如果我等待它完全成熟，这本书就永远不会完成。

我并不期待你能够全面掌握 Julia，或成为一个成熟的数据科学家。如果你渴望扩展技能，学习解决老问题的新方法，并严格按照本书的进度进行学习，那么 Julia 就会成为你进行数据分析的一个有效工具。

目录