

<<集体智慧编程>>

图书基本信息

书名：<<集体智慧编程>>

13位ISBN编号：9787121075391

10位ISBN编号：7121075393

出版时间：2009-1

出版时间：电子工业出版社

作者：TOBY SEGARAN

页数：334

译者：莫映,王开福

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：<http://www.tushu007.com>

## 前言

无论是有意还是无意，越来越多投身于互联网的人们已经制造出了相当多的数据，这给了我们无数潜在的机会来洞悉用户体验、商业营销、个人偏好和通常所谓的人类行为（human behavior）。本书向大家介绍了一个新兴的领域，称为聚集型智慧（collective intelligence）。这一领域涵盖了诸多方法，借助这些方法我们可以从众多Web站点处（这些站点的名字或许你曾经有所耳闻）提取到值得关注的重要数据；借助这些方法我们还可以从使用自己应用程序的用户那里搜集信息，并对我们所掌握的数据进行分析和理解。

本书的目的是要带领你超越以数据库为后端的简单应用系统，并告诉你如何利用自己和他人每天搜集到的信息来编写更为智能的程序。

**先决条件 Prerequisites** 本书的代码示例是用Python语言编写的，因此熟悉Python编程将会有助于你对算法的理解，不过笔者给出了所有算法的解释说明，所以其他语言的程序员也能看懂。对于已经了解了像Ruby或Perl这样高级语言的程序员，Python代码应该是非常容易理解的。本书的目的不是要作为一本学习编程的指导书，因此尤为重要的一点在于，为了熟悉基本概念，我们最好已编写过足够多的代码才行。

如果懂得递归和一点点函数式编程（functional programming）的基本概念，那么我们会发觉书中的内容是很容易理解的。

本书并不假设你已经具备了任何有关数据分析、机器学习或统计学方面的知识。笔者在尝试以尽可能浅显易懂的方式来解释数学概念，不过具备一点三角学和统计学的基本知识将会对你理解算法有所助益。

## &lt;&lt;集体智慧编程&gt;&gt;

## 内容概要

《集体智慧编程》( Programming Collective Intelligence Building Smart Web 2.0 Applications ) 以机器学习与计算统计为主题背景, 专门讲述如何挖掘和分析Web上的数据和资源, 如何分析用户体验、市场营销、个人品味等诸多信息, 并得出有用的结论, 通过复杂的算法来从Web网站获取、收集并分析用户的数据和反馈信息, 以便创造新的用户价值和商业价值。

全书内容翔实, 包括协作过滤技术( 实现关联产品推荐功能 )、集群数据分析( 在大规模数据集中发掘相似的数据子集 )、搜索引擎核心技术( 爬虫、索引、查询引擎、PageRank算法等 )、搜索海量信息并进行分析统计得出结论的优化算法、贝叶斯过滤技术( 垃圾邮件过滤、文本过滤 )、用决策树技术实现预测和决策建模功能、社交网络的信息匹配技术、机器学习和人工智能应用等。

《集体智慧编程》( Programming Collective Intelligence Building Smart Web 2.0 Applications ) 向读者介绍了: 令在线零售商向用户提供商品或媒体推荐的协作型过滤技术; 在一个大型数据集中检测相似项群组的聚类方法; 在针对某一问题的数以百万计的可能题解中进行搜索, 并从中选出最优解的优化算法; 用于垃圾过滤技术的贝叶斯过滤器, 如何根据单词类型及其他特征对文档进行分类; 用于对在线约会站点的用户进行配对的支持向量机; 用于问题求解的智能进化技术——随着玩游戏的次数逐渐增多, 计算机玩家如何通过改进自身代码的方式来发展技能。

《集体智慧编程》( Programming Collective Intelligence Building Smart Web 2.0 Applications ) 的每一章后都有练习, 这些练习对算法进行了扩展, 使其变得更加强大。

让我们超越以数据库为后端的简单应用系统, 挖掘互联网数据的价值, 为我所用!

深入讲解蕴藏于智能Web 2.0应用背后的算法原理, 介绍挖掘和分析用户贡献的内容, 以创造新的用户价值和商业价值。

## &lt;&lt;集体智慧编程&gt;&gt;

## 书籍目录

前言第1章 集体智慧引言什么是集体智慧什么是机器学习机器学习的局限真实生活中的例子学习型算法的其他用途第2章 提供推荐协作型过滤搜集偏好寻找相近的用户推荐物品匹配商品构建一个基于del.icio.us的链接推荐系统基于物品的过滤使用MovieLens数据集基于用户进行过滤还是基于物品进行过滤练习第3章 发现群组监督学习和无监督学习单词向量分级聚类绘制树状图列聚类K-均值聚类针对偏好的聚类以二维形式展现数据有关聚类的其他事宜练习第4章 搜索与排名搜索引擎的组成一个简单的爬虫程序建立索引查询基于内容的排名利用外部回指链接从点击行为中学习练习第5章 优化组团旅游描述题解成本函数随机搜索爬山法模拟退火算法遗传算法真实的航班搜索涉及偏好的优化网络可视化其他可能的应用场合练习第6章 文档过滤过滤垃圾信息文档和单词对分类器进行训练计算概率朴素分类器费舍尔方法将经过训练的分类器持久化过滤博客订阅源对特征检测的改进使用Akismet替代方法练习第7章 决策树建模预测注册用户引入决策树对树进行训练选择最合适的拆分方案以递归方式构造树决策树的显示对新的观测数据进行分类决策树的剪枝处理缺失数据处理数值型结果对住房价格进行建模对“热度”评价进行建模什么时候使用决策树练习第8章 构建价格模型构造一个样本数据集k-最近邻算法为近邻分配权重交叉验证不同类型的变量对缩放结果进行优化不对称分布使用真实数据——eBay API何时使用k-最近邻算法练习第9章 高阶分类：核方法与SVM婚介数据集数据中的难点基本的线性分类分类特征对数据进行缩放处理理解核方法支持向量机使用LIBSVM基于Facebook的匹配练习第10章 寻找独立特征搜集一组新闻先前的方法非负矩阵因式分解结果呈现利用股票市场的数据练习第11章 智能进化什么是遗传编程将程序以树形方式表示构造初始种群测试题解对程序进行变异交叉构筑环境一个简单的游戏更多可能性练习第12章 算法总结贝叶斯分类器决策树分类器神经网络支持向量机k-最近邻聚类多维缩放非负矩阵因式分解优化附录A：第三方函数库附录B：数学公式索引

<<集体智慧编程>>

媒体关注与评论

“太棒了！”

对于初学这些算法的开发者而言，我想不出有比这本书更好的选择了，而对于像我这样学过A J的老朽而言，我也想不出还有什么更好的办法能够让自己重温这些知识的细节。

”——Dan Russell，资深技术经理，Google “Toby的这本书非常成功地将机器学习算法这一复杂的议题拆分成了一个个既实用又易懂的例子，我们可以直接利用这些例子来分析当前网络上的社会化交互作用。

假如我早两年读过这本书，就会省去许多宝贵的时间，也不至于走那么多的弯路了。

”——Tim Wolters.CTO.Collective Intellect

## <<集体智慧编程>>

### 编辑推荐

想了解蕴藏在搜索排名、商品推荐、社会化书签以及在线婚介应用背后的巨大威力吗？

《集体智慧编程》（Programming Collective Intelligence Building Smart Web 2.0 Applications）的内容引人入胜，它将会告诉我们如何构造Web 2.0应用，使其能够挖掘有大量用户参与的互联网应用所产生的海量数据。

利用书中介绍的这些复杂算法，可以编写出智能程序、访问其他Web站点的数据集、从我们自己的应用程序中搜集用户数据，进而分析和理解这些数据。

《集体智慧编程》（Programming Collective Intelligence Building Smart Web 2.0 Applications）将引领我们进入机器学习与计算统计的世界，并解释如何得出有关用户体验、市场营销、个人品味以及我们和他人每天搜集的用户行为方面的结论。

书中对每一个算法都进行了详细的描述，并附以简洁的代码，这些代码可以直接用于我们的web站点、博客、维基，或者其他特定的应用。

《集体智慧编程》（Programming Collective Intelligence Building Smart Web 2.0 Applications）是Web开发者、架构师、应用工程师等的绝佳选择。

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:<http://www.tushu007.com>