

《社交网站的数据挖掘与分析(第2版)》

图书基本信息

书名：《社交网站的数据挖掘与分析(第2版)》

13位ISBN编号：9787111486994

出版时间：2015-1-28

作者：Matthew A. Russell

页数：370

译者：苏统华,魏通,赵逸雪,王烁行,刘智月

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.tushu000.com

《社交网站的数据挖掘与分析(第2版)》

内容概要

社交网站数据如同深埋地下的“金矿”，如何利用这些数据来发现哪些人正通过社交媒介进行联系？他们正在谈论什么？或者他们在哪儿？本书第2版对上一版内容进行了全面更新和修订，它将揭示回答这些问题的方法与技巧。你将学到如何获取、分析和汇总散落于社交网站（包括Facebook、Twitter、LinkedIn、Google+、GitHub、邮件、网站和博客等）的数据，以及如何通过可视化找到你一直在社交世界中寻找的内容和你闻所未闻的有用信息。

借助IPython Notebook、自然语言工具包、NetworkX和其他科学计算工具挖掘主流社交网站

使用高级文本挖掘技术（如聚类和TF-IDF）来提取人类语言数据中有价值的知识

通过发现GitHub上人、编程语言和代码工程间的亲密性，构建兴趣图谱

利用D3.js进行交互式可视化，充分发挥HTML5和JavaScript工具包的灵活特性

以“问题 - 解决方案 - 讨论”的方式详细讲解深入挖掘Twitter数据的实用技术，并提供代码示例

《社交网站的数据挖掘与分析（原书第2版）》的配套代码在公开的GitHub代码库中进行维护，可以通过一站式虚拟机来访问，你只需要使用方便易用的IPython Notebook，即可进入愉快的交互式学习情景。

《社交网站的数据挖掘与分析(第2版)》

作者简介

Matthew A. Russell Digital Reasoning Systems公司首席技术官（CTO）、Zaffra公司负责人。作为一名计算机科学家，他热衷于数据挖掘、开源软件开发和创造技术以扩展人类智能。

苏统华，博士，硕士生导师，CUDA研究中心以及教学中心负责人。主要研究方向包括：物联网大数据智能信息处理、大规模并行计算、模式识别、智能媒体交互与计算等。作为自然手写中文文本识别的开拓者，四年内代表工作被同行大篇幅他引约300次；他所建立的HIT-MW库为全世界100多家科研院所采用；目前负责国家自然科学基金项目2项。2013年，他领导的研究组在文档分析和识别国际会议（ICDAR ' 2013）上获得手写汉字识别竞赛的双料冠军；2014年，两项手写文字识别核心技术授权给某高新技术公司，正在为超过200万终端用户提供技术服务。著有英文专著《Chinese Handwriting Recognition: An Algorithmic Perspective》（德国施普林格出版社），出版5本大数据分析方面的译作（机械工业出版社）。

书籍目录

《社交网站的数据挖掘与分析（原书第2版）》	
前言	
1	
第一部分 社交网络导引	
序幕	
13	
第1章 挖掘Twitter：探索热门话题、发现人们的谈论内容等	
15	
1.1 概述	
15	
1.2 Twitter风靡一时的原因	
16	
1.3 探索Twitter API	
18	
1.4 分析140字的推文	
33	
1.5 本章小结	
47	
1.6 推荐练习	
48	
1.7 在线资源	
48	
第2章 挖掘Facebook：分析粉丝页面、查看好友关系等	
50	
2.1 概述	
51	
2.2 探索Facebook的社交图谱API	
51	
2.3 分析社交图谱联系	
62	
2.4 本章小结	
85	
2.5 推荐练习	
86	
2.6 在线资源	
86	
第3章 挖掘LinkedIn：分组职位、聚类同行等	
88	
3.1 概述	
89	
3.2 探索LinkedIn API	
89	
3.3 数据聚类速成	
94	
3.4 本章小结	
124	
3.5 推荐练习	

125	
3.6 在线资源	126
第4章 挖掘Google+：计算文档相似度、提取搭配等	127
4.1 概述	128
4.2 探索Google+ API	128
4.3 TF-IDF简介	138
4.4 用TF-IDF查询人类语言数据	145
4.5 本章小结	164
4.6 推荐练习	165
4.7 在线资源	165
第5章 挖掘网页：使用自然语言处理理解人类语言、总结博客内容等	167
5.1 概述	168
5.2 抓取、解析、爬取网页	168
5.3 通过解码语法来探索语义	174
5.4 以实体为中心的分析：范式转换	192
5.5 人类语言数据处理分析的质量	200
5.6 本章小结	203
5.7 推荐练习	203
5.8 在线资源	204
第6章 挖掘邮箱：分析谁和谁说什么以及说的频率等	206
6.1 概述	207
6.2 获取和处理邮件语料库	207
6.3 分析Enron语料库	225
6.4 探索和可视化时序趋势	241
6.5 分析你自己的邮件数据	244

6.6 本章小结	250
6.7 推荐练习	251
6.8 在线资源	251
第7章 挖掘GitHub：检查软件协同习惯、构建兴趣图谱等	253
7.1 概述	254
7.2 探索GitHub的API	254
7.3 使用属性图为数据建模	260
7.4 分析GitHub兴趣图谱	264
7.5 本章小结	286
7.6 推荐练习	287
7.7 在线资源	287
第8章 挖掘带标记语义网：提取微格式、推断资源描述框架等	289
8.1 概述	290
8.2 微格式：易于实现的元数据	290
8.3 从语义标记过渡到语义网：一个小插曲	304
8.4 语义网：发展中的变革	304
8.5 本章小结	310
8.6 推荐的练习	311
8.7 在线资源	311
第二部分 Twitter实用指南	
第9章 Twitter实用指南	317
9.1 访问Twitter的API（开发目的）	318
9.2 使用OAuth访问Twitter的API（产品目的）	319
9.3 探索流行话题	323
9.4 查找推文	324

9.5 构造方便的函数调用	325
9.6 使用文本文件存储JSON数据	326
9.7 使用MongoDB存储和访问JSON数据	327
9.8 使用信息流API对Twitter数据管道抽样	329
9.9 采集时序数据	330
9.10 提取推文实体	332
9.11 特定的推文范围内查找最流行的推文	333
9.12 特定的推文范围内查找最流行的推文实体	335
9.13 对频率分析制表	336
9.14 查找转推了状态的用户	337
9.15 提取转推的属性	339
9.16 创建健壮的Twitter请求	340
9.17 获取用户个人资料信息	343
9.18 从任意的文本中提取推文实体	344
9.19 获得用户所有的好友和关注者	345
9.20 分析用户的好友和关注者	347
9.21 获取用户的推文	348
9.22 爬取好友关系图	350
9.23 分析推文内容	351
9.24 提取链接目标摘要	353
9.25 分析用户收藏的推文	356
9.26 本章小结	357
9.27 推荐练习	358
9.28 在线资源	359
第三部分 附录	

附录A 关于本书虚拟机体验的信息

363

附录B OAuth入门

364

附录C Python和IPython Notebook的使用技巧

368

《社交网站的数据挖掘与分析(第2版)》

精彩短评

- 1、社交网站数据挖掘入门书，主要讲一些网站API和python工具包的使用示例
- 2、一本讲技术的书，好多代码，快速略过.....作为文科生仅仅来看看思路
- 3、主要是介绍国外几个社交网站的python 访问API，鲜有有意思的数据挖掘案例和思路，且代码也有点旧了

《社交网站的数据挖掘与分析(第2版)》

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.tushu000.com