

<<智能Web算法>>

图书基本信息

书名：<<智能Web算法>>

13位ISBN编号：9787121139192

10位ISBN编号：7121139197

出版时间：2011-11

出版时间：电子工业出版社

作者：Haralambos Marmanis,Dmitry Babenko

页数：400

译者：阿稳,陈钢

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>

<<智能Web算法>>

内容概要

本书涵盖了五类重要的智能算法：搜索、推荐、聚类、分类和分类器组合，并结合具体的案例讨论了它们在Web应用中的角色及要注意的问题。除了第1章的概要性介绍以及第7章对所有技术的整合应用外，第2~6章以代码示例的形式分别对这五类算法进行了介绍。

<<智能Web算法>>

作者简介

Haralambos (Babis) Marmanis

博士是一位把机器学习技术应用于工业界的先行者，也是供应链管理的世界级专家。

Dmitry

Babenko曾经为银行、保险、供应链管理与商务智能公司设计过应用与基础架构。

<<智能Web算法>>

书籍目录

1 什么是智能Web ?

1.1 智能Web应用实例

1.2 智能应用的基本要素

1.3 什么应用会受益于智能?

1.3.1 社交网络

1.3.2 Mashup

1.3.3 门户网站

1.3.4 维基

1.3.5 文件分享网站

1.3.6 网络游戏

1.4 如何构建智能应用 ?

1.4.1 检查功能和数据

1.4.2 获取更多的数据

1.5 机器学习、数据挖掘及其他

1.6 智能应用中八个常见的误区

1.6.1 误区1：数据是可靠的

1.6.2 误区2：计算能马上完成

1.6.3 误区3：不用考虑数据规模

1.6.4 误区4：不考虑解决方案的可扩展性

1.6.5 误区5：随处使用同样的方法

1.6.6 误区6：总是能知道计算时间

1.6.7 误区7：复杂的模型更好

1.6.8 误区8：存在无偏见的模型

1.7 小结

1.8 参考资料

2 搜索

2.1 用Lucene实现搜索

2.1.1 理解Lucene代码

2.1.2 搜索的基本步骤

2.2 为什么搜索不仅仅是索引 ?

2.3 用链接分析改进搜索结果

2.3.1 PageRank简介

2.3.2 计算PageRank向量

2.3.3 alpha：网页间跳转的影响

2.3.4 理解幂方法

2.3.5 结合索引分值和PageRank分值

2.4 根据用户点击改进搜索结果

2.4.1 用户点击初探

2.4.2 朴素贝叶斯分类器的使用

2.4.3 整合Lucene索引、PageRank和用户点击

2.5 Word、PDF等无链接文档的排序

2.5.1 DocRank算法简介

<<智能Web算法>>

- 2.5.2 DocRank的原理
- 2.6 大规模实现的有关问题
- 2.7 用户得到了想要的结果吗？
- 精确度和查全率
- 2.8 总结
- 2.9 To Do
- 2.10 参考资料
- 3 推荐系统
 - 3.1 一个在线音乐商店：基本概念
 - 3.1.1 距离与相似度的概念
 - 3.1.2 走近相似度的计算
 - 3.1.3 什么才是最好的相似度计算公式？
 - 3.2 推荐引擎是怎么工作的
 - 3.2.1 基于相似用户的推荐
 - 3.2.2 基于相似条目的推荐
 - 3.2.3 基于内容的推荐
 - 3.3 推荐朋友、文章与新闻报道
 - 3.3.1 MyDiggSpace.com简介
 - 3.3.2 发现朋友
 - 3.3.3 DiggDelphi的内部工作机制
 - 3.4 像Netflix.com那样推荐电影
 - 3.4.1 电影数据集的介绍及推荐器
 - 3.4.2 数据标准化与相关系数
 - 3.5 大规模的实现与评估
 - 3.6 总结
 - 3.7 To Do
 - 3.8 参考资料
- 4 聚类：事物的分组
 - 4.1 聚类的需求
 - 4.1.1 网站中的用户组：案例研究
 - 4.1.2 用SQL order by子句分组
 - 4.1.3 用数组排序分组
 - 4.2 聚类算法概述
 - 4.2.1 基于分组结构的聚类算法分类
 - 4.2.2 基于数据类型和结构的聚类算法分类
 - 4.2.3 根据数据规模的聚类算法分类
 - 4.3 基于链接的算法
 - 4.3.1 树状图：基本的聚类数据结构
 - 4.3.2 基于链接的算法概况
 - 4.3.3 单链接算法
 - 4.3.4 平均链接算法
 - 4.3.5 最小生成树算法
 - 4.4 k-means算法
 - 4.4.1 初识k-means算法
 - 4.4.2 k-means的内部原理
 - 4.5 鲁棒的链接型聚类（ROCK）

<<智能Web算法>>

- 4.5.1 ROCK简介
- 4.5.2 为什么ROCK这么强大？
- 4.6 DBSCAN
 - 4.6.1 基于密度的算法简介
 - 4.6.2 DBSCAN的原理
- 4.7 超大规模数据聚类
 - 4.7.1 计算复杂性
 - 4.7.2 高维度
- 4.8 总结
- 4.9 To Do
- 4.10 参考资料
- 5 分类：把事物放到它该在的地方
 - 5.1 对分类的需求
 - 5.2 分类器的概述
 - 5.2.1 结构分类算法
 - 5.2.2 统计分类算法
 - 5.2.3 分类器的生命周期
 - 5.3 邮件的自动归类与垃圾邮件过滤
 - 5.3.1 朴素贝叶斯分类
 - 5.3.2 基于规则的分类
 - 5.4 用神经网络做欺诈检测
 - 5.4.1 交易数据中关于欺诈检测的一个用例
 - 5.4.2 神经网络概览
 - 5.4.3 一个可用的神经网络欺诈检测器
 - 5.4.4 神经网络欺诈检测器剖析
 - 5.4.5 创建通用神经网络的基类
 - 5.5 你的结果可信吗？
 - 5.6 大数据集的分类
 - 5.7 总结
 - 5.8 To Do
 - 5.9 参考资料
- 6 分类器组合
 - 6.1 信贷价值：分类器组合案例研究
 - 6.1.1 数据的简要说明
 - 6.1.2 为真实问题生成人工数据
 - 6.2 用单分类器做信用评估
 - 6.2.1 朴素贝叶斯的基准线
 - 6.2.2 决策树基准线
 - 6.2.3 神经网络基线
 - 6.3 在同一个数据集中比较多个分类器
 - 6.3.1 McNemar检验
 - 6.3.2 差额比例检验
 - 6.3.3 Cochran Q检验与F检验
 - 6.4 Bagging: bootstrap聚合 (bootstrap aggregating)
 - 6.4.1 bagging实例

<<智能Web算法>>

- 6.4.2 bagging分类器底层细节
- 6.4.3 分类器集成
- 6.5 Boosting：一种迭代提高的方法
 - 6.5.1 boosting分类器实例
 - 6.5.2 boosting分类器底层细节
- 6.6 总结
- 6.7 To Do
- 6.8 参考资料
- 7 智能技术大汇集：一个智能新闻门户
 - 7.1 功能概览
 - 7.2 获取并清洗内容
 - 7.2.1 各就位、预备、开抓！
 - 7.2.2 搜索预备知识回顾
 - 7.2.3 一个抓取并处理好的新闻数据集
 - 7.3 搜索新闻
 - 7.4 分配新闻类别
 - 7.4.1 顺序问题
 - 7.4.2 使用NewsProcessor类进行分类
 - 7.4.3 分类器
 - 7.4.4 分类策略：超越底层的分类
 - 7.5 用NewsProcessor类创建新闻分组
 - 7.5.1 聚类全部文章
 - 7.5.2 在一个新闻类别中聚类文章
 - 7.6 基于用户评分的动态内容展示
 - 7.7 总结
 - 7.8 To Do
 - 7.9 参考资料
- 附录A BeanShell简介
 - A.1 什么是BeanShell？
 - A.2 为什么使用BeanShell？
 - A.3 运行BeanShell
 - A.4 参考资料
- 附录B 网络采集
 - B.1 爬虫组件概况
 - B.1.1 采集的步骤
 - B.1.2 我们的简单爬虫
 - B.1.3 开源Web爬虫
 - B.2 参考资料
- 附录C 数学知识回顾
 - C.1 向量和矩阵
 - C.2 距离的度量
 - C.3 高级矩阵方法
 - C.4 参考资料
- 附录D 自然语言处理

<<智能Web算法>>

D.1 参考资料
附录E 神经网络
E.1 参考资料
索引

<<智能Web算法>>

编辑推荐

算法是解决问题的一系列步骤。

为实现有价值的Web应用（如推荐引擎、智能化搜索、内容组织系统等），本书提供了清晰的、精心组织过的算法模式。

利用这些技术，你可以捕获用户原始而重要的信息，并把它们应用于实践中以获取相应的收益。

用户数据中包含大量有价值的关联信息，它们往往无法通过人工观察而直观地获取，对于希望从这些数据中挖掘信息的Web开发者来说，玛若曼尼斯、巴宾寇编著的《智能Web算法》是一本很好的手册。

作者作为一名Web开发者，拥有丰富的实践经验，加上多年来对机器学习领域技术的专研，使得本书对技术的解释清晰明了，读者可快速将其用于解决自己的问题。

同时，本书提供的Java程序展示了如何搭建一个智能的应用，以及如何从用户的行为中进行学习，这是一笔现成的财富。

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>