第一图书网, tushu007.com

<<推荐系统实践>>

图书基本信息

书名：<<推荐系统实践>>

13位ISBN编号：9787115281586

10位ISBN编号：7115281580

出版时间：2012-6

出版时间：人民邮电出版社

作者：项　亮

页数：197

字数：319000

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

说起本书，还要追溯到2010年3月份的ResysChina推荐系统大会。
在那次会议上，我遇到了刘江老师。
刘老师看过我之前写的一些推荐系统方面的博客，希望我能总结总结，写本简单的书。
当时国内还没有推荐系统方面的书，而国外已经有这方面的专业书了，因此图灵公司很想出版一本介绍推荐系统的书。
所以，去年7月博士毕业时，我感觉有时间可以总结一下这方面的工作了，于是准备开始写这本书。
写这本书的目的有下面几个。
首先，从个人角度讲，虽然写博士论文时已经总结了读博期间在推荐系统方面的工作，但并没有全部涉及整个推荐系统的各个方面，因此我很希望通过写作这本书全面地阅读一下相关的文献，并在此基础上总结一下推荐系统各个方面的发展现状，供大家参考。
其次，最近几年从事推荐系统研究的人越来越多，这些人中有些原来是工程师，对机器学习和数据挖掘不太了解，有些是在校学生，虽然对数据挖掘和机器学习有所了解，却对业界如何实现推荐系统不太清楚。
因此，我希望能够通过本书让工程师了解推荐系统的相关算法，让学生了解如何将自己了解的算法实现到一个真实的工业系统中去。
一般认为，推荐系统这个研究领域源于协同过滤算法的提出。
这么说来，推荐系统诞生快20年了。
这期间，很多学者和公司对推荐系统的发展起到了重要的推动作用，各种各样的推荐算法也层出不穷。
本书希望将这20年间诞生的典型方法进行总结。
但由于方法太多，这些方法的归类有很多不同的方式。
比如，可以按照数据分成协同过滤、内容过滤、社会化过滤，也可以按照算法分成基于邻域的算法、基于图的算法、基于矩阵分解或者概率模型的算法。
为了方便读者入门，本书基本采用数据分类的方法，每一章都介绍了一种可以用于推荐系统设计的、新类型的用户数据，然后介绍如何通过各种方法利用该数据，最后在公开数据集上评测这些方法。
当然，不是所有数据都有公开的数据集，并且不是所有算法都可以进行离线评测。
因此，在遇到没有数据集或无法进行离线评测的问题时，本书引用了一些著名学者的实验结果来说明各种方法的效果。
为了使本书同时适合工程师和在校学生阅读，本书在写作中同时使用了两种介绍方法。
一种是利用公式，这样方便有一些理论基础的同学很快明白算法的含义。
另一种是利用代码，这样可以方便工程师迅速了解算法的含义。
不过因为本人是学生出身，工程经验还不是特别足，所以有些代码写得不是那么完美，还请工程师们海涵。
本书一开始写的时候有3位作者，除了我之外还有豆瓣的陈义和腾讯的王益。
他们两位都是这方面的前辈，在写作过程中提出了很多宝贵的意见。
但因为二位工作实在太繁忙，所以本书主要由我操刀。
但书中的很多论述融合了大家的思想和经验，是我们很多次讨论的结果。
因此在这里感谢王益和陈义二位合作者，虽然二位没有动笔，但对这本书做出了很大的贡献。
其次，还要感谢吴军老师和谷文栋为本书作序。
感谢谷文栋、稳国柱、张夏天各自审阅了书中部分内容，提出了很多宝贵的意见。
感谢我在Hulu的同事郑华和李航，郑华给了我充分的时间完成这本书，对这本书能够按时出版功不可没，而李航审阅了书中的部分内容，提出了很多有价值的修改意见。
最后感谢我的父母和妻子，他们在我写作过程中给予了很大照顾，感谢他们的辛勤付出。

内容概要

《推荐系统实践》通过大量代码和图表全面系统地阐述了和推荐系统有关的理论基础，介绍了评价推荐系统优劣的各种标准(比如覆盖率、满意度)和方法(比如AB
测试)，总结了当今互联网领域中各种和推荐有关的产品和服务。
另外，本书为有兴趣开发推荐系统的读者给出了设计和实现推荐系统的方法与技巧，并解答了在真实场景中应用推荐技术时最常遇到的一些问题。

《推荐系统实践》适合对推荐技术感兴趣的读者学习参考。

作者简介

项亮
毕业于中国科学技术大学和中国科学院自动化所，研究方向为机器学习和推荐系统，现任职于北京Hulu软件技术开发有限公司，从事视频推荐的研究和开发。
2009年参加Netflix
Prize推荐系统比赛获得团体第二名，且于当年参与创建了Resys China推荐系统社区。

<<推荐系统实践>>

书籍目录

第1章　好的推荐系统　1
1.1　什么是推荐系统　1
1.2　个性化推荐系统的应用　4
1.2.1　电子商务　4
1.2.2　电影和视频网站　8
1.2.3　个性化音乐网络电台　10
1.2.4　社交网络　12
1.2.5　个性化阅读　15
1.2.6　基于位置的服务　16
1.2.7　个性化邮件　17
1.2.8　个性化广告　18
1.3　推荐系统评测　19
1.3.1　推荐系统实验方法　20
1.3.2　评测指标　23
1.3.3　评测维度　34

第2章　利用用户行为数据　35
2.1　用户行为数据简介　36
2.2　用户行为分析　39
2.2.1　用户活跃度和物品流行度的分布　39
2.2.2　用户活跃度和物品流行度的关系　41
2.3　实验设计和算法评测　41
2.3.1　数据集　42
2.3.2　实验设计　42
2.3.3　评测指标　42
2.4　基于邻域的算法　44
2.4.1　基于用户的协同过滤算法　44
2.4.2　基于物品的协同过滤算法　51
2.4.3　UserCF和ItemCF的综合比较　59
2.5　隐语义模型　64
2.5.1　基础算法　64
2.5.2　基于LFM的实际系统的例子　70
2.5.3　LFM和基于邻域的方法的比较　72
2.6　基于图的模型　73
2.6.1　用户行为数据的二分图表示　73
2.6.2　基于图的推荐算法　73

第3章　推荐系统冷启动问题　78
3.1　冷启动问题简介　78
3.2　利用用户注册信息　79
3.3　选择合适的物品启动用户的兴趣　85
3.4　利用物品的内容信息　89
3.5　发挥专家的作用　94

第4章　利用用户标签数据　96

4.1　UGC标签系统的代表应用　97
4.1.1　Delicious　97
4.1.2　CiteULike　98
4.1.3　Last.fm　98
4.1.4　豆瓣　99
4.1.5　Hulu　99
4.2　标签系统中的推荐问题　100
4.2.1　用户为什么进行标注　100
4.2.2　用户如何打标签　101
4.2.3　用户打什么样的标签　102
4.3　基于标签的推荐系统　103
4.3.1　实验设置　104
4.3.2　一个最简单的算法　105
4.3.3　算法的改进　107
4.3.4　基于图的推荐算法　110
4.3.5　基于标签的推荐解释　112
4.4　给用户推荐标签　115
4.4.1　为什么要给用户推荐标签　115
4.4.2　如何给用户推荐标签　115
4.4.3　实验设置　116
4.4.4　基于图的标签推荐算法　119
4.5　扩展阅读　119

第5章　利用上下文信息　121
5.1　时间上下文信息　122
5.1.1　时间效应简介　122
5.1.2　时间效应举例　123
5.1.3　系统时间特性的分析　125
5.1.4　推荐系统的实时性　127
5.1.5　推荐算法的时间多样性　128
5.1.6　时间上下文推荐算法　130
5.1.7　时间段图模型　134
5.1.8　离线实验　136
5.2　地点上下文信息　139
5.3　扩展阅读　143

第6章　利用社交网络数据　144
6.1　获取社交网络数据的途径　144
6.1.1　电子邮件　145
6.1.2　用户注册信息　146
6.1.3　用户的位置数据　146
6.1.4　论坛和讨论组　146
6.1.5　即时聊天工具　147
6.1.6　社交网站　147
6.2　社交网络数据简介　148
　　社交网络数据中的长尾分布　149
6.3　基于社交网络的推荐　150

6.3.1　基于邻域的社会化推荐算法　151
6.3.2　基于图的社会化推荐算法　152
6.3.3　实际系统中的社会化推荐算法　153
6.3.4　社会化推荐系统和协同过滤推荐系统　155
6.3.5　信息流推荐　156
6.4　给用户推荐好友　159
6.4.1　基于内容的匹配　161
6.4.2　基于共同兴趣的好友推荐　161
6.4.3　基于社交网络图的好友推荐　161
6.4.4　基于用户调查的好友推荐算法对比　164
6.5　扩展阅读　165

第7章　推荐系统实例　166
7.1　外围架构　166
7.2　推荐系统架构　167
7.3　推荐引擎的架构　171
7.3.1　生成用户特征向量　172
7.3.2　特征-物品相关推荐　173
7.3.3　过滤模块　174
7.3.4　排名模块　174
7.4　扩展阅读　178

第8章　评分预测问题　179
8.1　离线实验方法　180
8.2　评分预测算法　180
8.2.1　平均值　180
8.2.2　基于邻域的方法　184
8.2.3　隐语义模型与矩阵分解模型　186
8.2.4　加入时间信息　192
8.2.5　模型融合　193
8.2.6　Netflix Prize的相关实验结果　195

后记　196

章节摘录

版权页： 插图： 众所周知，为了解决信息过载的问题，已经有无数科学家和工程师提出了很多天才的解决方案，其中代表性的解决方案是分类目录和搜索引擎。
而这两种解决方案分别催生了互联网领域的两家著名公司——雅虎和谷歌。
著名的互联网公司雅虎凭借分类目录起家，而现在比较著名的分类目录网站还有国外的DMOZ、国内的Hao123等。
这些目录将著名的网站分门别类，从而方便用户根据类别查找网站。
但是随着互联网规模的不断扩大，分类目录网站也只能覆盖少量的热门网站，越来越不能满足用户的需求。
因此，搜索引擎诞生了。
以谷歌为代表的搜索引擎可以让用户通过搜索关键词找到自己需要的信息。
但是，搜索引擎需要用户主动提供准确的关键词来寻找信息，因此不能解决用户的很多其他需求，比如当用户无法找到准确描述自己需求的关键词时，搜索引擎就无能为力了。
和搜索引擎一样，推荐系统也是一种帮助用户快速发现有用信息的工具。
和搜索引擎不同的是，推荐系统不需要用户提供明确的需求，而是通过分析用户的历史行为给用户的兴趣建模，从而主动给用户推荐能够满足他们兴趣和需求的信息。
因此，从某种意义上说，推荐系统和搜索引擎对于用户来说是两个互补的工具。
搜索引擎满足了用户有明确目的时的主动查找需求，而推荐系统能够在用户没有明确目的的时候帮助他们发现感兴趣的新内容。
从物品的角度出发，推荐系统可以更好地发掘物品的长尾（long tail）。
美国《连线》杂志主编Chris Anderson在2004年发表了“ The Long Tail ”（长尾）一文并于2006年出版了《长尾理论》一书。
该书指出，传统的80／20原则（80%的销售额来自于20%的热门品牌）在互联网的加入下会受到挑战。
互联网条件下，由于货架成本极端低廉，电子商务网站往往能出售比传统零售店更多的商品。
虽然这些商品绝大多数都不热门，但与传统零售业相比，这些不热门的商品数量极其庞大，因此这些长尾商品的总销售额将是一个不可小觑的数字，也许会超过热门商品（即主流商品）带来的销售额。
主流商品往往代表了绝大多数用户的需求，而长尾商品往往代表了一小部分用户的个性化需求。
因此，如果要通过发掘长尾提高销售额，就必须充分研究用户的兴趣，而这正是个性化推荐系统主要解决的问题。
推荐系统通过发掘用户的行为，找到用户的个性化需求，从而将长尾商品准确地推荐给需要它的用户，帮助用户发现那些他们感兴趣但很难发现的商品。
要了解推荐系统是如何工作的，可以先回顾一下现实社会中用户面对很多选择时做决定的过程。
仍然以看电影为例，一般来说，我们可能用如下方式决定最终看什么电影。
向朋友咨询。
我们也许会打开聊天工具，找几个经常看电影的好朋友，问问他们有没有什么电影可以推荐。
甚至，我们可以打开微博，发表一句“ 我要看电影 ”，然后等待热心人推荐电影。
这种方式在推荐系统中称为社会化推荐（social recommendation），即让好友给自己推荐物品。
我们一般都有喜欢的演员和导演，有些人可能会打开搜索引擎，输入自己喜欢的演员名，然后看看返回结果中还有什么电影是自己没有看过的。
比如我非常喜欢周星驰的电影，于是就去豆瓣搜索周星驰，发现他早年的一部电影我还没看过，于是就会看一看。
这种方式是寻找和自己之前看过的电影在内容上相似的电影。
推荐系统可以将上述过程自动化，通过分析用户曾经看过的电影找到用户喜欢的演员和导演，然后给用户推荐这些演员或者导演的其他电影。
这种推荐方式在推荐系统中称为基于内容的推荐（content-based filtering）。

我们还可能查看排行榜，比如著名的IMDB电影排行榜，看看别人都在看什么电影，别人都喜欢什么电影，然后找一部广受好评的电影观看。
这种方式可以进一步扩展：如果能找到和自己历史兴趣相似的一群用户，看看他们最近在看什么电影，那么结果可能比宽泛的热门排行榜更能符合自己的兴趣。
这种方式称为基于协同过滤（collaborative filtering）的推荐。

媒体关注与评论

“工程师大都喜欢‘In Action’型的书籍，但这并非‘又一本’工具型的‘In Action’读物，透过它你将涉足现代互联网公司孜孜以求的用户核心价值所在——个性化服务。
也许一次全新的旅程就从这里开始。
”　　——阿稳，豆瓣资深算法工程师　　“作者结合了多年的推荐系统理论研究和在Hulu的具体实践经验，汲取精华并以深入浅出的方式展示给读者。
无论是对于刚入门的新手还是推荐领域的老兵，这本书都是不可多得的必备参考，在此我诚挚地向大家推荐它。
”　　——郑华，Hulu资深软件开发主管　　“2009年8月，我和项亮一起发起了Resys China——一个面向推荐系统领域的专业社区。
在组织Resys China业内分享活动的过程中，我们迫切感受到，出版《推荐系统实践》这样一本传授实战经验的书籍，对推动这个领域的发展是多么必要。
项亮作为国内推荐系统领域一位理论与实践并重的专家，把最具实用价值的推荐技术进行了系统整理，深入浅出地呈现到读者面前。
作为一本主要面向业内人员的技术书籍，这点尤其难能可贵。
个性化推荐技术是最具人文关怀的技术之一，它尊重个体，相信每个人都是与众不同的，在这个以‘人’为中心的社会化时代，它的兴起与发扬光大只是时间问题。
我与项亮相识，是因为对推荐技术的热爱，希望借助此书，可以让更多的人成为朋友。
”　　——谷文栋，个性化推荐社区Resys China发起人　　“从大家经常使用的相关搜索、话题推荐、电子商务的各种产品推荐，到社交网络上的交友推荐等，推荐系统在今天互联网的产品和应用中被广泛采用。
但是，至今还没有一本书系统地从理论上对此进行分析和论述。
《推荐系统实践》恰恰弥补了这个空白。
”　　——吴军，腾讯副总裁，《数学之美》和《浪潮之巅》作者

编辑推荐

《推荐系统实践》适合对推荐技术感兴趣的读者学习参考。

名人推荐

“工程师大都喜欢‘In Action’型的书籍，但这并非‘又一本’工具型的‘In Action’读物，透过它你将涉足现代互联网公司孜孜以求的用户核心价值所在——个性化服务。
也许一次全新的旅程就从这里开始。
”——阿稳，豆瓣资深算法工程师“作者结合了多年的推荐系统理论研究和在Hulu的具体实践经验，汲取精华并以深入浅出的方式展示给读者。
无论是对于刚入门的新手还是推荐领域的老兵，这本书都是不可多得的必备参考，在此我诚挚地向大家推荐它。
”——郑华，Hulu资深软件开发主管“2009年8月，我和项亮一起发起了Resys China——一个面向推荐系统领域的专业社区。
在组织Resys China业内分享活动的过程中，我们迫切感受到，出版《推荐系统实践》这样一本传授实战经验的书籍，对推动这个领域的发展是多么必要。
项亮作为国内推荐系统领域一位理论与实践并重的专家，把最具实用价值的推荐技术进行了系统整理，深入浅出地呈现到读者面前。
作为一本主要面向业内人员的技术书籍，这点尤其难能可贵。
个性化推荐技术是最具人文关怀的技术之一，它尊重个体，相信每个人都是与众不同的，在这个以‘人’为中心的社会化时代，它的兴起与发扬光大只是时间问题。
我与项亮相识，是因为对推荐技术的热爱，希望借助此书，可以让更多的人成为朋友。
”——谷文栋，个性化推荐社区Resys China发起人“从大家经常使用的相关搜索、话题推荐、电子商务的各种产品推荐，到社交网络上的交友推荐等，推荐系统在今天互联网的产品和应用中被广泛采用。
但是。
至今还没有一本书系统地从理论上对此进行分析和论述。
《推荐系统实践》恰恰弥补了这个空白。
”——吴军，腾讯副总裁，《数学之美》和《浪潮之巅》作者

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