《机器学习》

图书基本信息

书名：《机器学习》

13位ISBN编号：9787111417316

10位ISBN编号：7111417313

出版时间：2013-4-1

出版社：机械工业出版社

作者：（美）Drew Conway,John Myles White

页数：320

译者：陈开江,刘逸哲,孟晓楠,罗森林 审校

《机器学习》

前言

【译者序】当今各行业，尤其是互联网，数据规模越来越大，要从中有效地发现模式来提高生产力，用传统的方式已经几乎不可能，只能借助计算机来完成诸多使命。因此，机器学习这一新兴的学科变得越来越重要，它已经在搜索、推荐、数据挖掘等多个领域闪耀光芒。机器学习是一门交叉学科，内容涉及概率论、统计学、高等数学、计算机科学等多门学科。该学科致力于设计一种让计算机具有“学习”能力的算法，通过发现经验数据中隐藏的模式，实现对未知数据的预测。大数据时代是机器学习最美好的时代，因为数据不再是问题，各类问题都可以收集到海量的数据。但是，对于很多人来说，这一门交叉学科本身却神秘而陌生，对于没有系统学习过相关基础学科的人来说尤其感到“高不可攀”。如今已出版的机器学习相关书籍中，很多都有这个特点：公式多，晦涩难懂。这让很多程序员出身的人望而却步。然而，在第一次读到本书的英文版时，译者就彻底相信：机器学习完全可以讲解得通俗易懂，让知识的传递实现“润物细无声”。本书秉承的原则是：实践出真知，只要多动手，没有攻克不了的技术难题。因此作者预期的阅读对象是如电脑黑客般的人，要求对技术有发自内心的求知欲和好奇心，愿意自己动手而非纸上谈兵。全书精心选择了12个机器学习案例，由浅入深，面面俱到，既有基础知识（如数据分析），也有当前热门的社交网站推荐案例。书中的每一个案例都由作者娓娓道来，逐一剖析关键算法的代码，没有丝毫学究气息，触动每个机器学习初学者的内心最深处。书中所有算法都采用R语言实现。R语言是一门用于统计学的开源脚本语言，基于它的开源性，有来自世界各地的开源拥护者贡献的各种统计学相关的程序包，稳定且方便，尤其是它对数据可视化的支持，更是一柄利器，既轻巧又实用。书中所有源代码和数据在原书的官方网站上都可以免费下载。在阅读过程中，犹如作者亲至身侧，为你讲解代码和思路，为你排除错误和优化效果。全书案例既有分类问题，也有回归问题；既包含监督学习，也涵盖无监督学习。所选择的案例妙趣横生，如分析UFO目击记录、破译密码、预测股票、分析美国参议员“结党”的情况，等等，这里就不“剧透”了，大家自己去享受学习的乐趣吧。书中12个案例之间的依赖关系不是特别强（除R语言基础知识外，其余某几章仅有个别知识点之间存在依赖性），可以像连续剧一样，逐一播放，也可以像一个个小品一般，挑感兴趣的内容分别播放。学习完这些案例之后，相信你会窥见机器学习的一斑，然后再根据自己的实际情况更深入地学习。本书翻译工作由三位来自互联网世界的工程师通力协作完成，其中，来自新浪微博的陈开江负责完成前言及第1～4章的翻译；来自阿里B2B的刘逸哲负责完成第5、8、9和11章的翻译；来自阿里一淘的孟晓楠负责完成第6、7、10和12章的翻译；同时，全书审校工作由来自北京理工大学的罗森林教授义务承担。本书能够得以出版，首先要感谢机械工业出版社的吴怡编辑，是她给了我们三位工程师这个学习知识并传递知识的机会，她经验丰富，在翻译过程中给予了我们许多建设性的指导意见。其次，要感谢罗森林教授，他在百忙之中为我们担任全书的审校工作，从而让国内的机器学习者能感受到这本书应有的魅力。最后，我们要感谢互联网，因为译者与本书的缘分始于互联网，从看到原书、报名翻译、组成翻译团队、翻译过程中的讨论，所有这样都是通过互联网完成的。虽然经过罗森林教授认真审校并且给我们提出了宝贵意见，但是由于译者本身水平有限，书中译文势必还存在不妥甚至错误之处，恳请机器学习界的广大前辈、同仁们不吝赐教，促使我们继续为大家更好地传递先进技术，让更多机器学习爱好者成为机器学习的黑客。我们坚信集体智慧是再高的个人智慧都无法企及的，因此真诚希望大家一起来贡献自己的智慧。无论是对翻译本身有任何意见或建议，还是对机器学习方面有心得，都欢迎大家到我们的微博上交流、切磋，我们一起贡献自己的智慧，在集体智慧中互相学习，共同进步。

《机器学习》

内容概要

《机器学习》

作者简介

【作者介绍】
Drew Conway 机器学习专家，拥有丰富的数据分析与处理工作经验。目前主要利用数学、统计学和计算机技术研究国际关系、冲突和恐怖主义等。他曾作为研究员在美国情报和国防部门供职数年。他拥有纽约大学政治系博士学位，曾为多种杂志撰写文章，是机器学习领域的著名学者。
John Myles White 机器学习专家，拥有丰富的数据分析与处理工作经验。目前主要从理论和实验的角度来研究人类如何做出决定，同时还是几个流行的R语言程序包的主要维护者，包括ProjectTemplate和log4r。他拥有普林斯顿大学哲学系博士学位，曾为多家技术杂志撰稿，发表过许多关于机器学习的论文，并在众多国际会议上发表演讲。
【译者介绍】
罗森林 博士，教授，博导。现任北京理工大学信息系统及安全对抗实验中心主任、专业责任教授。国防科技工业局科学技术委员会成员；《中国医学影像技术杂志》、《中国介入影像与治疗学》编委会委员；全国大学生信息安全技术专题邀请赛专家组副组长；中国人工智能学会智能信息安全专业委员会委员等。主要研究方向为信息安全、数据挖掘、媒体计算、中文信息处理等。负责或参加完成国家自然科学基金、国家科技支撑计划、863计划、国家242计划等省部级以上项目40余项。已发表学术论文90余篇，出版著作8部，出版译著1部，获授权专利3项。
陈开江　新浪微博搜索部研发工程师，曾独立负责微博内容反垃圾系统、微博精选内容挖掘算法、自助客服系统（包括自动回复、主动挖掘、舆情监测）等项目，目前主要从事社交挖掘、推荐算法研究、机器学习、自然语言处理相关工作，研究兴趣是社交网络的个性化推荐。
刘逸哲　阿里巴巴，CBU基础平台部搜索与推荐团队核心技术与query分析方向负责人，机器学习技术领域及圈子负责人。曾任中国雅虎相关性团队、自然语言处理团队算法工程师；AvePoint.inc开发工程师，从事企业级搜索引擎开发。研究兴趣是机器学习、自然语言处理及个性化推荐等算法在大规模数据上的应用。
孟晓楠　一淘广告技术，阿里非搜索广告算法负责人，负责用户行为分析、建模与细分，RTB竞价算法，展示广告CTR预估与SEM优化。曾工作于网易杭州研究院，参与过分布式全文检索系统和网易博客产品的数据挖掘算法开发。研究兴趣是计算广告技术、机器学习、大数据技术、信息检索等。

《机器学习》

书籍目录

章节摘录

版权页： 插图： 在图11—7中，我们专注于网络的左半部分，并且把边都删除了，以便于更容易观察节点的标签。快速浏览一遍这部分聚类的Twitter用户名，很明显Drew的这部分网络包含了Drew在Twitter上关注的数据专家。首先我们看到的是知名的数据专家，比如浅绿色的蒂姆·奥莱利（timoreilly）和Nathan Yau（flowingdata），因为他们都是自成一体的。紫色和红色的组也很有趣，因为它们都含有数据黑客，但是被一个关键因子分成两部分：Drew的紫色好友都是数据圈子的杰出成员，例如：HilaryMason（hmason）、PeteSkomoroch（peteskomoroch）、Jake Hofman（jakehofman），但是他们没有一位是R语言圈子的活跃成员。另一方面，红色的节点都是R语言圈子的活跃成员，包括HadlevWickham（hadleywickham）、David Smith（revodavid）、Gary King（kinggary）。此外，力导向算法成功地把这些圈子成员放到一起，并且把属于这两个圈子的那些节点放到圈子的边缘。我们可以看到John（johnmyleswhite）是紫色的，但是他被放到很多红色节点中间。这是因为John在这两个圈子中都是杰出成员，而且数据也反映了这一点。其他的这类例子包括：JD Long（cmastication）和Josh Reich（i2pi）。 尽管Drew花了很长时间和数据圈子成员交流（包括R用户和非R用户数据圈子成员），但是Drew也使用Twitter与满足其他兴趣的圈子交流。其中一个特别的兴趣是他的学术职业生涯，他关注国家安全技术和政策。在图11—8中，我们突出了Drew网络的右半部分，它包含了来自这些兴趣相关的圈子的成员。和数据专家组类似，这部分包含了2个子组，一个是蓝色的，另外一个是绿色的。和前面的例子一样，节点的分割颜色和摆放位置可以反映出他们在网络中扮演的角色。 蓝色分割中的Twitter用户铺得很开：一部分离Drew很近，在网络的左边，而另外一些在网络的右边，接近绿色的组。那些靠近左边的用户与技术在国家安全中的角色这一话题有关，这些用户包括：Sean Gourley（sgourley）、Lewis Shepherd（1ewisshepherd）和Jeffrey Carr（Jeffrey Carr）。那些靠近绿色组的用户更加关注国家安全政策，和绿色组中的成员相似。在绿色组中，我们看到很多Twitter上著名的国家安全圈子成员，包括：AndrewExum（abumuqawama）、Joshua Foust（joshua Foust）和Daveed Gartenstein—Ross（daveedgr）。和前面一样，有趣的是，那些属于两个组的人被放置到聚类边缘，例如：Chris Albon（chrisalbon），他在两个圈子中都很杰出。

媒体关注与评论

“O’ReillyRadar博客有口皆碑。”——Wired“O’Reilly凭借一系列（真希望当初我也想到了）非凡想法建立了数百万美元的业务。”——Business2.0“O’ReillyConference是聚集关键思想领袖的绝对典范。”——CRN“一本O’Reilly的书就代表一个有用、有前途、需要学习的主题。”——IrishTimes“Tim是位特立独行的商人，他不光放眼于最长远、最广阔的视野并且切实地按照YogiBerra的建议去做了：‘如果你在路上遇到岔路口，走小路（岔路）。’回顾过去Tim似乎每一次都选择了小路，而且有几次都是一闪即逝的机会，尽管大路也不错。”——LinuxJournal

《机器学习》

编辑推荐

《机器学习:实用案例解析》编辑推荐：1.《机器学习:实用案例解析》是机器学习和数据挖掘领域的经典图书，基础理论与实践完美的结合，是一部逻辑紧密、内容详实，适合所有相关技术人员的参考书。2.《机器学习:实用案例解析》两名作者都具有丰富的数据分析、处理工作经验，是机器学习实践技术方面的积极实践者。

《机器学习》

名人推荐

O ’ Reilly Radar博客有口皆碑。——WiredO ’ Reilly凭借一系列（真希望当初我也想到了）非凡想法建立了数百万美元的业务。——Business 2.0O ’ Reilly Conference是聚集关键思想领袖的绝对典范。——CRN一本O ’ Reilly的书就代表一个有用、有前途、需要学习的主题。——Irish TimesTim是位特立独行的商人，他不光放眼于最长远、最广阔的视野并且切实地按照Yogi Berra的建议去做了：‘ 如果你在路上遇到岔路口，走小路（岔路）。 ’ 回顾过去Tim似乎每一次都选择了小路，而且有几次都是一闪即逝的机会，尽管大路也不错。——Linux Journal

《机器学习》

精彩短评

1、这本书真是骗钱的厉害，重新排个版，去掉废话，能减少三分之一的页数。
2、翻译的不算完美，有的地方读起来蛮吃力，总的来说不错。
3、因为人需要洞悉多因素的相关性，所以教会机器干最无聊的收集（行为数据）、分析、（初步）结论。因此，这里汉语“学习”应当理解为“记住”并“认识到”。
4、用R做机器学习，这种手把手做案例的书就很好啊。
5、很好！
6、随书的程序和数据没有地方下载，不利于学习
7、:
TP181/0250
8、非常不错，推荐。实用举例，不蔓不枝。
9、需要这样的案例，毕竟刚刚接触
10、写给统计学家的机器学习书，写给MLer的统计分析书，写给R语言初学者的实践进阶书，写给开发工程师的算法入门书。这本书把所有的公式都忽略掉了，比大名鼎鼎的集体智慧编程还要夸张和简单....
11、标题说是机器学习-实用案例解析，而实际上原标题是Machine Learning for Hacker。内容基本是实践，点到为止，不深入，较泛。另外翻译质量也不太满意
12、举一反三！！再来两遍~
13、应该叫 machine learning for statistics idiots
14、基本上把代码试了一遍，虽然英文名字里面有个hacker，但里面讲的东西倒是step by step ,适合入门，里面对机器学习讲的不多，主要讲了回归，分类，聚类，最后还捎着讲了SVM，基于R的实操，亲手试一下，有好处的。
15、了解个大概，原理解释的比较少（不过可能也不用），唯一就是对R无感。
16、与集体智慧编程是一类的书
17、还是比较适合研发同学
18、很不错也很实用，可以入手。送的速度很快 第二天就到了！
19、非常好的机器学习入门课程，以案例讲解算法，分析深入详细，对于希望学习该课程的人员来说，非常值得阅读。中文版的我已经买了。。。
20、内容有一定价值，但是实在不实用
21、借助R语言，把机器学习算法都变成黑盒，专注于数据的整理和思路。角度蛮特别的一本书，对实干比较有启发意义。
22、这本书讲的比较基础，但是很清晰。我正在读，会很认真的读。
23、怎么没有案例中的数据下载~~！！没数据怎么分析？！不知道你们好评的评的是啥。。
24、案例很丰富，需要下力气来研究
25、讲究实用，这是最好的，使用才是极客的风范，否则就只能算是学院
26、都是浅出了，没有深入的感觉，看的时候还是必要有理论参考书比较好
27、我自己参与翻译的，不敢给最高分。但是，这本书的确是好，就是好，好好好！哈哈哈~~~
28、很精致，内容很有趣，正在学习
29、机器学习已经比较困难了，这个有范例总体来说比较方便但是范例的完整度不够，对于初学者而言最好还是“手把手”“一步一步”这样的组织形式可以节省大量的构建环境的时间
30、前半部分讲的非常精彩，适合入门人员，很多原理娓娓道来，虽然看不懂R，但是对于入门人员来讲，是需要了解原理的。后半部分很多理论就很复杂，又是一笔带过，就显得粗糙。继续深入挖掘吧。
31、。。。
32、准备先看完斯坦福的机器学习再看这本书，大致翻了下，感觉只是结构很好，先给4星
33、hackers，算法实践
34、蹭风口。
35、看了一点。感觉很多地方翻译的不妥。还是看英文原版吧。

36、机器学习的常用场景讲解
37、比较适合初学者。
38、总而言之，这本书作为机器学习的入门教材还是可以的。但是要注意两点，书中所阐述的机器学习算法和概念是相对最为基础的，尽管很简单，但是已经包含了最重要最常用的思想，很多内容举重若轻，新人看和高手看绝对会获得不同的理解；另一点，书中前两章基本反应了全书中要求的R语言水平，作为复习和练习都是不错的实践，但是作为R语言的入门就有些不合适了。
39、基本没怎么看懂，
40、R确实比较强大。
41、实在是太浅显了
42、<<Machine Learning For Hackers>> 适合初学者实践。特别最后一章的Twitter关系网可视化，酷炫！
43、想学习机器学习的，Oreilly的书都很不错。
44、难度中，有很多好思路
45、作者不懂数学，只是懂R
46、书名应该改，改成　R语言统计学的实用案例解析
47、从基础开始一步一步教的，R语言使用者而非编写者所写，角度有所不同。
48、当小说书看即可，当r语言的机器学习包工具书看也可。
49、呜呜，没看懂，但是不明觉厉。。。

《机器学习》

精彩书评

1、对于机器学习，一直困惑于缺乏实践，缺少可操作的入手点。也一直在读理论理论，有种总是在打敲边鼓的感觉。本书举了不少例子，基于R语言的，终于看到一些实际的例子了。或许以后可以找出其中一个例子进行学习。总体来说，这本书还行，还是有可读性的。
2、对于机器学习，一直困惑于缺乏实践，缺少可操作的入手点。也一直在读理论理论，有种总是在打敲边鼓的感觉。本书举了不少例子，基于R语言的，终于看到一些实操的例子了。或许以后可以找出其中一个例子进行学习。总体来说，这本书还行，还是有可读性的。

《机器学习》

版权说明