

预测性文本挖掘基础 [Fundamentals of Predictive Text Mining]



[预测性文本挖掘基础 \[Fundamentals of Predictive Text Mining\]_下载链接1](#)

著者:[美] 绍洛姆·韦斯（Weiss S.） 著，赵仲孟，侯迪 译

[预测性文本挖掘基础 \[Fundamentals of Predictive Text Mining\]_下载链接1](#)

标签

评论

很不错的专业书籍，很棒，值得推荐！

活动的时候下单的，送来有点晚，不过能理解。

书不错，可学习的东西还有很多

买来学习学习买来学习学习

不错????????????

书的质量杠杠的，不过从北京寄过来，快递时间慢很多。

很不错，书的内容很好，很喜欢！赞一个！

老公给单位买的，京东买书很方便。

很好，下次还会来买，

给别人买的 还不错

书我还没有仔细阅读，但整体感觉不错，纸质和印刷质量也很好，应该是正版图书，值得一读，相信会对自己有很大帮助！只是读英文教材多少还是有些吃力！
京东服务也很好，送货速度也比较快，而且这次用了满200减100的券，价格也非常实惠，希望京东将来多做优惠！

送货及时，包装很好，购买方便，下次还来。

很好，买得非常值，会继续买。

用来入门,慢慢看吧.

正在看，还不错还不错很基础

等读了再来评价，但愿我记得

呵呵，有点上当的感觉

对自己的学习希望有帮助

图书的质量很好，都是正品书，价格也比较优惠。具体的内容还没有看过，书是帮别人带的，同事们看了之后评价还都蛮高的。京东的书发货快，基本第一天下单，第二天就能到了。客服和物流的服务态度也都挺不错的，有耐心。

到货很快，京东物流赞一个

很高深的书，翻译的嘛，我也没啥好说了。

很专业的书，用了统计学基础之后再看看吧

本书内容易懂，章节分明

帮学生买的，希望有用吧

预测性文本挖掘的好用工具，好用的教程

浅显易懂，不错，值得购买

好书，正版，很实用，值得购买

还没看感觉可以，希望有打印的书单和发票钉在一起。

东西很不错 用起来很方便 点个赞

书里有很多实例，但可能还是偏广泛

教材不错！用心看完再说

不错的书，学习用，不错的书，学习用

趁打折多买点趁打折多买点趁打折多买点

该书很好，内容不错,文本挖掘

到货及时，不知写得怎样，等看完后再发表意见！

还没用呢呀，有前瞻性

送货很快，书不错。。

中文的文本挖掘教材太少了，这本翻译的国外书籍弥补了一些空缺。这本书从书名可以看出是讲基础知识，而且是理论为主，看完这本书再去国外原版的高级书籍应该更为轻松。

非常满意，五星

做论文要用的，买来看看，学习学习。

便宜，送货快，还会再买。。。

整体感觉不错，快递很给力

文本中存在无限的挖掘价值，因为他们是作者内心真实的反映。

内容到位哦

很好的书

挺好的！

货不错.....

像盗版的。。。

用得上

书还行

不错哦

不少人都买这本书了。虽然出版时间较早，但可做基础参考书。

可以。

说实话，用例很一般，也许是因为2012年出版的，内容有些陈旧。

大学毕业后，我先后在加利福尼亚和中国大陆的Intel公司工作。最初，我打算工作两年之后回学校读研究生，但是幸福时光飞逝而过，转眼就过去了六年。那时，我意识到我必须回到校园。我不想上夜校或进行在线学习，我就想坐在大学校园里吸纳学校传授的所有知识。在大学里，最好的方面不是你研修的课程或从事的研究，而是一些外围活动：与人会面、参加研讨会、加入组织、旁听课程，以及学习未知的知识。

在2008年，我帮助筹备一个招聘会。我同一个大金融机构的人交谈，他们希望我去应聘他们机构的一个对信用卡建模（判断某人是否会偿还贷款）的岗位。他们问我对随机分析了解多少，那时，我并不能确定“随机”一词的意思。他们提出的工作地点令我无法接受，所以我决定不再考虑了。但是，他们说的“随机”让我很感兴趣，于是我拿来课程目录，寻找含有“随机”字样的课程，我看到了“离散随机系统”。我没有注册就直接旁听了这门课，完成课后作业，参加考试，最终被授课教授发现。但是她很仁慈，让我继续学习，这让我非常感激。上这门课，是我第一次看到将概率应用到算法中。在这之前，我见过一些算法将平均值作为外部输入，但这次不同，方差和均值都是这些算法中的内部值。这门课主要讨论时间序列数据，其中每一段数据都是一个均匀间隔样本。我还找到了名称中包含“机器学习”的另一门课程。该课程中的数据并不假设满足时间的均匀间隔分布，它包含更多的算法，但严谨性有所降低。再后来我意识到，在经济系、电子工程系和计算机科学系的课程中都会讲授类似的算法。

2009年初，我顺利毕业，并在硅谷谋得了一份软件咨询的工作。接下来的两年，我先后在涉及不同技术的八家公司工作，发现了最终构成这本书主题的两种趋势：第一，为了开发出竞争力强的应用，不能仅仅连接数据源，而需要做更多事情；第二，用人单位希望员工既懂理论也能编程。

程序员的大部分工作可以类比于连接管道，所不同的是，程序员连接的是数据流，这也为人们带了巨大的财富。举一个例子，我们要开发一个在线出售商品的应用，其中主要部分是允许用户来发布商品并浏览其他人发布的商品。为此，我们需要建立一个Web表单，允许用户输入所售商品的信息，然后将该信息传到一个数据存储器。要让用户看到其他用户所售商品的信息，就要从数据存储器获取这些数据并适当地显示出来。我可以确信，人们会通过这种方式挣钱，但是如果让要应用更好，需要加入一些智能因素。这些智能因素包括自动删除不适当的发布信息、检测不正当交易、给出用户可能喜欢的商品以及预测网站的流量等。为了实现这些目标，我们需要应用机器学习方法。对于最终用户而言，他们并不了解幕后的“魔法”，他们关心的是应用能有效运行，这也是好产品的标志。

一个机构会雇用一些理论家（思考者）以及一些做实际工作的人（执行者）。前者可能会将大部分时间花在学术工作上，他们的日常工作就是基于论文产生思路，然后通过高级工具或数学进行建模。后者则通过编写代码与真实世界交互，处理非理想世界中的瑕疵，比如崩溃的机器或者带噪声的数据。完全区分这两类人并不是个好想法，很多成功的机构都认识到这一点。（精益生产的一个原则就是，思考者应该自己动手去做实际工作。）当招聘经费有限时，谁更能得到工作，思考者还是执行者？很可能是执行者，但是现实中用人单位希望两种人都要。很多事情都需要做，但当应用需要更高要求的算法时，那么需要的人员就必须能够阅读论文，领会论文思路并通过代码实现，如此反复下去。

在这之前，我没有看到在机器学习算法方面缩小思考者和执行者之间差距的书籍。本书的目的就是填补这个空白，同时介绍机器学习算法的使用，使得读者能够构建更成功的应用。

从手足无措，到剑气凛然，再到收束锋芒，柴静自言，这是一条逐渐放下自己的过程，在卢安克身上，她“体会到了一些从没想过、未曾明白的东西——把自我交付出去，从

此活在命运之中的必然与自由”。事实上，放下自我和寻找自我，殊途同归。最初，柴静对这本书最大的犹豫，在于太多的“我”是否不妥。老六说，关键不在“我”，而在要写“人”。这本书中，是给柴静本人留下强烈生命印象的人，她说，这些人构成了当下的“我”，因此这本看上去主题并不鲜明的书，实际上却有着一个最为核心的话题，报道自己。

说到这里，我突然想起蒙田，数百年前的一位智者，他在《随想录》里毫不嫌疑，大谈自己，开卷即说：“吾书素材无他，即吾人也”（je suis moi-meme la matiere de mon livre）。但就是这部“只写自己”的书，获得无数的隔代知音，不在于他的教谕，而在于他对自由的追求、对怀疑的执着。晚生一百多年的孟德斯鸠说，“在大多数作品中，我看到了写书的人；而在这一本书中，我却看到了一个思想者。”

伪善的思想家，思想不过名词而已；真正的思想者，思想是一个需要身体力行的动词，执行这个动作很艰难，“就像叶子从痛苦的蜷缩中要用力舒展一样，人也要从不假思索的蒙昧里挣脱”。这个过程，只有也必须自己来完成，只有每天坚持精神拉练，思想的肌肉才不会松懈。

柴静说自己是个丢三落四的人，一下飞机，电脑就不见了，好友范铭说，柴静是她见过唯一一个手机到处放，而且从不设密码的人。就这么个人。白天在新闻里播“两会”，不小心出溜一个“万人空巷”的词儿，回家能把自己狠狠地K了一通，日记还要记上一笔，“我跟你讲柴小静，就冲你今天赶这么用这个成语，将来杀人放火的事儿也干得出来”。

她挑剔、警戒、为难、批判的对象只有一个，就是自己，她小心翼翼地，把汹涌的感性之河导入理性的河床中，无时无刻不盯着自己，就像盯着另一个人，就像蒙田说的那样，“我要不停地向我自己走去，因为我要持续不断地指责自己。”

柴静反复说，这本书中没有什么想法是人们不熟悉的，或者异乎寻常的，其中都是普通的、有关人性的东西。1784年康德写下《什么是启蒙》，两百年后，福柯再次就同一命题写下文章，但在康德之前，启蒙就是个问题，福柯之后也依然还是，只要人类尚存，和人本身密切相关的话题就不会结束，自由也是。谁都知道，在充斥着暴力、专横的动荡不安中

[预测性文本挖掘基础 \[Fundamentals of Predictive Text Mining\]_下载链接1](#)

书评

[预测性文本挖掘基础 \[Fundamentals of Predictive Text Mining\]_下载链接1](#)