* 出版社： [人民邮电出版社](https://book.jd.com/publish/%E4%BA%BA%E6%B0%91%E9%82%AE%E7%94%B5%E5%87%BA%E7%89%88%E7%A4%BE_1.html)
* ISBN：9787115373557
* 版次：2
* 商品编码：11572052
* 品牌：[异步图书](https://www.jd.com/pinpai/1-1713-266205.html)
* 包装：平装
* 开本：16开
* 出版时间：2014-11-01
* 用纸：胶版纸
* 页数：312
* 正文语种：中文

**编辑推荐**

　　《数学之美》第一版荣获国家图书馆第八届文津图书奖；   
　　2014中国书业评选年度图书；央视新闻推荐的学科敲门砖；   
　　入选新闻出版广电总局“2014年向全国青少年推荐百种优秀图书书目”；   
　　荣获2012-2013年度全行业畅销书；   
　　业界专家和高校教师推荐：信息领域大学生必读好书。   
　　《浪潮之巅》、《文明之光》作者吴军博士新力作，李开复作序推荐，Google黑板报百万点击！   
　　第二版增加了大数据和机器智能等新内容，并根据专家和读者反馈做了修订，与时俱进更新了部分内容。

**内容简介**

　　几年前，“数学之美”系列文章原刊载于谷歌黑板报，获得上百万次点击，得到读者高度评价。正式出版前，吴军博士几乎把所有文章都重写了一遍，为的是把高深的数学原理讲得更加通俗易懂，让非专业读者也能领略数学的魅力。   
　　《数学之美》上市后深受广大读者欢迎，并荣获国家图书馆第八届文津图书奖。读者说，读了《数学之美》，才发现大学时学的数学知识，比如马尔科夫链、矩阵计算，甚至余弦函数原来都如此亲切，并且栩栩如生，才发现自然语言和信息处理这么有趣。   
　　而今，数学在信息产业中的应用越来越广泛，因此，作者在第二版中增加了一些内容，尤其是针对大数据和机器学习的内容，以便满足人们对当下技术的学习需求。

**作者简介**

　　吴军，博士，毕业于清华大学和美国约翰·霍普金斯大学，是自然语言处理和搜索专家，硅谷风险投资人。获奖畅销书《浪潮之巅》及《数学之美》的作者。  
　　吴军博士是谷歌公司早期员工之一。在谷歌，他和辛格（美国工程院院士，世界搜索专家）、Matt Cutts（谷歌反作弊官方发言人）等三位同事一起开创了网络搜索反作弊的研究领域，并因此获得谷歌工程奖。2003年，他和谷歌全球架构的总工程师朱会灿博士等共同成立了中日韩文搜索部门。吴军博士是当前谷歌中日韩文搜索算法的主要设计者。在谷歌期间，他还领导了许多研发项目，得到了当时公司首席执行官埃里克?施密特和创始人谢尔盖·布林的高度评价。  
　　2010年—2012年，他加盟腾讯公司，出任负责搜索和搜索广告的副总裁，同时担任国家重大专项“新一代搜索引擎和浏览器”项目的总负责人。2012年回到谷歌，负责开发了被认为是“下一代搜索”的谷歌自动问答系统。同年，他作为创始合伙人共同创立了中关村硅谷风险投资基金（ZPark Venture）。  
　　吴军博士在国内外发表过数十篇论文，曾获得全国人机语音智能接口会议的论文奖和Eurospeech的论文奖。他还获得了十余项美国和国际专利。  
　　吴军博士还担任约翰·霍普金斯大学工学院董事会董事和校国际事务委员会顾问，他也长期担任中国工业和信息化部的专家顾问。同时，他也是数家投资基金、创业公司的董事和顾问。

**精彩书评**

　　★我大学的专业是计算数学，但读到吴军老师的“数学之美”系列文章，才发现马尔科夫过程、矩阵计算，甚至余弦函数原来都如此亲切，并且栩栩如生；才发现自然语言和信息处理这么有趣；才真正明白“数学是科学的皇后”这句名言。相信认真读完这本《数学之美》的朋友们，算法功力都会暴涨N倍，更重要的是发现了数学背后的无穷魅力，学会欣赏数学之美。 

　　——蒋涛（CSDN《程序员》创始）

　　★看到《数学之美》，是谷歌黑板报上的连载文章。里面的公式并不是很多，但是很多看似颇为复杂的概念，吴军老师却能够如讲故事般娓娓道出，着实看出作者对这些问题有着深入且独到的见解，读后受益匪浅。这次有幸在《数学之美》出版之前拜读了初稿，欣喜看到新书在章节连贯和语言方面都较黑板报的连载文章有了较大的提高，相信每一个喜欢数学、乐意欣赏数学之美的读者，一定会觉得开卷有益。 

　　——张磊（微软亚洲研究院主管研究员）

　　★我不做研究，也自觉没有做研究的底子。然而，数年前看到吴军老师的《数学之美》系列时仍然还是被深深地迷住了。正如作为一个十几年的科幻爱好者，深信在平凡的生活和工作之余应得闲仰望星空一样，作为生活在信息社会的个体，在上微博、搜Google、发邮件之余，关上显示器，能够透过《数学之美》这样的杰作，一窥纷繁涌动的数字世界背后的引擎数学之美，实乃一件幸事。 

　　——刘未鹏（《暗时间》作者）

　　★接触吴军老师的“数学之美”系列，是在搜索bloomfilter资料时，读了其中一篇后，就把其他的文章都读了，感触很多：首先，改变了观点：原以为在计算机系学到的数学基础在工作中一无是处，现在懂得：知识要落地，重要的是理解知识的由来；其次，任何复杂的问题可以用简单的方式去解决，我们往往会陷入不断给问题增加难度的复杂解法，而忽视了简单直接有效的方法。   
　　“数学之美”系列文章，整体和细节的度掌握得很好，通过具体的例子让读者学到的是思考问题的方式，同时留了很多问题给愿意钻研的人做进一步深入思考。BTW，“数学之美”系列，是我在技术领域介绍中读过的的文章之一，让人学会如何化繁为简，如何用数学去解决工程问题，如何跳出固有思维不断去思考创新。   
　　——岑文初（淘宝开放平台技术产品负责人）

**目录**

目录

i　　第一版读者赞誉   
iv　 第二版出版说明   
vii　第一版序言   
x　　第二版序言   
xii　第二版前言   
第1章　文字和语言 vs 数字和信息　1   
文字和语言与数学，从产生起原本就有相通性，虽然它们的发展一度分道扬镳，但是最终还是能走到一起。   
1　信息   
2　文字和数字   
3　文字和语言背后的数学   
4　小结   
第2章　自然语言处理——从规则到统计　15   
人类对机器理解自然语言的认识走了一条大弯路。早期的研究集中采用基于规则的方法，虽然解决了一些简单的问题，但是无法从根本上将自然语言理解实用化。直到20多年后，人们开始尝试用基于统计的方法进行自然语言处理，才有了突破性进展和实用的产品。   
1　机器智能   
2　从规则到统计   
3　小结   
第3章　统计语言模型　27   
统计语言模型是自然语言处理的基础，并且被广泛应用于机器翻译、语音识别、印刷体或手写体识别、拼写纠错、汉字输入和文献查询。   
1　用数学的方法描述语言规律   
2　延伸阅读：统计语言模型的工程诀窍   
3　小结   
第4章　谈谈分词　41   
中文分词是中文信息处理的基础，它同样走过了一段弯路，目前依靠统计语言模型已经基本解决了这个问题。   
1　中文分词方法的演变   
2　延伸阅读：如何衡量分词的结果   
3　小结   
第5章　隐含马尔可夫模型　50   
隐含马尔可夫模型最初应用于通信领域，继而推广到语音和语言处理中，成为连接自然语言处理和通信的桥梁。同时，隐含马尔可夫模型也是机器学习的主要工具之一。   
1　通信模型   
2　隐含马尔可夫模型   
3　延伸阅读：隐含马尔可夫模型的训练   
4　小结   
第6章　信息的度量和作用　60   
信息是可以量化度量的。信息熵不仅是对信息的量化度量，也是整个信息论的基础。它对于通信、数据压缩、自然语言处理都有很强的指导意义。   
1　信息熵   
2　信息的作用   
3　互信息   
4　延伸阅读：相对熵   
5　小结   
第7章　贾里尼克和现代语言处理　72   
作为现代自然语言处理的奠基者，贾里尼克教授成功地将数学原理应用于自然语言处理领域中，他的一生富于传奇色彩。   
1　早年生活   
2　从水门事件到莫妮卡·莱温斯基   
3　一位老人的奇迹   
第8章　简单之美——布尔代数和搜索引擎　82   
布尔代数虽然非常简单，却是计算机科学的基础，它不仅把逻辑和数学合二为一，而且给了我们一个全新的视角看待世界，开创了数字化时代。   
1　布尔代数   
2　索引   
3　小结   
第9章　图论和网络爬虫　89   
互联网搜索引擎在建立索引前需要用一个程序自动地将所有的网页下载到服务器上，这个程序称为网络爬虫，它的编写是基于离散数学中图论的原理。   
1　图论   
2　网络爬虫   
3　延伸阅读：图论的两点补充说明   
4　小结   
第10章　PageRank——Google的民主表决式网页排名技术　98   
网页排名技术PageRank是早期Google的杀手锏，它的出现使得网页搜索的质量上了一个大的台阶。它背后的原理是图论和线性代数的矩阵运算。   
1　PageRank 算法的原理   
2　延伸阅读：PageRank的计算方法   
3　小结   
第11章　如何确定网页和查询的相关性　104   
确定网页和查询的相关性是网页搜索的根本问题，其中确定查询中每个关键词的重要性有多高是关键。TF-IDF是目前通用的关键词重要性的度量，其背后的原理是信息论。   
1　搜索关键词权重的科学度量TF-IDF   
2　延伸阅读：TF-IDF的信息论依据   
3　小结   
第12章　有限状态机和动态规划——地图与本地搜索的核心技术　111   
地图和本地服务中要用到有限状态机和动态规划技术。这两项技术是机器智能和机器学习的工具，它们的应用非常广泛，还包括语音识别、拼写和语法纠错、拼音输入法、工业控制和生物的序列分析等。   
1　地址分析和有限状态机   
2　全球导航和动态规划   
3　延伸阅读：有限状态传感器   
4　小结   
第13章　Google AK-47的设计者——阿米特·辛格博士　121   
在所有轻武器中最有名的是AK-47冲锋枪，因为它从不卡壳，不易损坏，可在任何环境下使用，可靠性好，杀伤力大并且操作简单。Google的产品就是按照上述原   
则设计的。   
第14章　余弦定理和新闻的分类　127   
计算机虽然读不懂新闻，却可以准确地对新闻进行分类。其数学工具是看似毫不相干的余弦定理。   
1　新闻的特征向量   
2　向量距离的度量   
3　延伸阅读：计算向量余弦的技巧   
4　小结   
第15章　矩阵运算和文本处理中的两个分类问题　136   
无论是词汇的聚类还是文本的分类，都可以通过线性代数中矩阵的奇异值分解来进行。这样一来，自然语言处理的问题就变成了一个数学问题。   
1　文本和词汇的矩阵   
2　延伸阅读：奇异值分解的方法和应用场景   
3　小结   
第16章　信息指纹及其应用　142   
世间万物都有一个唯一标识的特征，信息也是如此。每一条信息都有它特定的指纹，通过这个指纹可以区别不同的信息。   
1　信息指纹   
2　信息指纹的用途   
3　延伸阅读：信息指纹的重复性和相似哈希   
4　小结   
第17章　由电视剧《暗算》所想到的——谈谈密码学的数学原理　153   
密码学的根本是信息论和数学。没有信息论指导的密码是非常容易被破解的。只有在信息论被广泛应用于密码学后，密码才真正变得安全。   
1　密码学的自发时代   
2　信息论时代的密码学   
3　小结   
第18章　闪光的不一定是金子——谈谈搜索引擎反作弊问题和搜索结果的权威性问题　162   
闪光的不一定是金子，搜索引擎中排名靠前的网页也未必是有用的网页。消除这些作弊网页的原理和通信中过滤噪音的原理相同。这说明信息处理和通信的很多原理是相通的。   
1　搜索引擎的反作弊   
2　搜索结果的权威性   
3　小结   
第19章　谈谈数学模型的重要性　171   
正确的数学模型在科学和工程中至关重要，而发现正确模型的途径常常是曲折的。正确的模型在形式上通常是简单的。   
第20章　不要把鸡蛋放到一个篮子里——谈谈最大熵模型　177   
最大熵模型是一个完美的数学模型。它可以将各种信息整合到一个统一的模型中，在信息处理和机器学习中有着广泛的应用。它在形式上非常简单、优美，而在实现时需要有精深的数学基础和高超的技巧。   
1　最大熵原理和最大熵模型   
2　延伸阅读：最大熵模型的训练   
3　小结   
第21章　拼音输入法的数学原理　186   
汉字的输入过程本身就是人和计算机之间的通信。好的输入法会自觉或不自觉地遵循通信的数学模型。当然要做出最有效的输入法，应当自觉使用信息论做指导。   
1　输入法与编码   
2　输入一个汉字需要敲多少个键——谈谈香农第一定理   
3　拼音转汉字的算法   
4　延伸阅读：个性化的语言模型   
5　小结   
第22章　自然语言处理的教父马库斯和他的优秀弟子们　197   
将自然语言处理从基于规则的研究方法转到基于统计的研究方法上，宾夕法尼亚大学的教授米奇马库斯功不可没。他创立了今天在学术界广泛使用的LCD语料库，同时培养了一大批精英人物。   
1　教父马库斯   
2　从宾夕法尼亚大学走出的精英们   
第23章　布隆过滤器　204   
日常生活中，经常要判断一个元素是否在一个集合中。布隆过滤器是计算机工程中解决这个问题最好的数学工具。   
1　布隆过滤器的原理   
2　延伸阅读：布隆过滤器的误识别问题   
3　小结   
第24章　马尔可夫链的扩展——贝叶斯网络　209   
贝叶斯网络是一个加权的有向图，是马尔可夫链的扩展。而从认识论的层面看：贝叶斯网络克服了马尔可夫链那种机械的线性约束，它可以把任何有关联的事件统一到它的框架下面。它在生物统计、图像处理、决策支持系统和博弈论中都有广泛的使用。   
1　贝叶斯网络   
2　贝叶斯网络在词分类中的应用   
3　延伸阅读：贝叶斯网络的训练   
4　小结   
第25章　条件随机场、文法分析及其他　217   
条件随机场是计算联合概率分布的有效模型，而句法分析似乎是英文课上英语老师教的东西，这两者有什么联系呢？   
1　文法分析——计算机算法的演变   
2　条件随机场   
3　条件随机场在其他领域的应用   
4　小结   
第26章　维特比和他的维特比算法　227   
维特比算法是现代数字通信中使用最频繁的算法，同时也是很多自然语言处理的解码算法。可以毫不夸张地讲，维特比是对我们今天生活的影响力最大的科学家之一，因为如今基于CDMA的3G移动通信标准主要就是他创办的高通公司制定的。   
1　维特比算法   
2　CDMA技术——3G移动通信的基础   
3　小结   
第27章　上帝的算法——期望最大化算法　238   
只要有一些训练数据，再定义一个最大化函数，采用EM算法，利用计算机经过若干次迭代，就可以得到所需要的模型。这实在是太美妙了，这也许是我们的造物主刻意安排的。所以我把它称作上帝的算法。   
1　文本的自收敛分类   
2　延伸阅读：期望最大化和收敛的必然性   
3　小结   
第28章　逻辑回归和搜索广告　244   
逻辑回归模型是一种将影响概率的不同因素结合在一起的指数模型，它不仅在搜索广告中起着重要的作用，而且被广泛应用于信息处理和生物统计中。   
1　搜索广告的发展   
2　逻辑回归模型   
3　小结   
第29章　各个击破算法和Google云计算的基础　249   
Google颇为神秘的云计算中最重要的MapReduce工具，其原理就是计算机算法中常用的“各个击破”算法，它的原理原来这么简单——将复杂的大问题分解成很多小问题分别求解，然后再把小问题的解合并成原始问题的解。由此可见，在生活中大量用到的、真正有用的方法常常都是简单朴实的。   
1　分治算法的原理   
2　从分治算法到MapReduce   
3　小结   
第30章　Google大脑和人工神经网络　254   
Google大脑并不是一个什么都能思考的大脑，而是一个很能计算的人工神经网络。因此，与其说Google大脑很聪明，不如说它很能算。不过，换个角度来说，随着计算能力的不断提高，计算量大但简单的数学方法有时能够解决很复杂的问题。   
1　人工神经网络   
2　训练人工神经网络   
3　人工神经网络与贝叶斯网络的关系   
4　延伸阅读：Google大脑   
5　小结   
第31章　大数据的威力——谈谈数据的重要性　273   
如果说在过去的40年里，主导全球IT产业发展的是摩尔定律，那么在今后的20年里，主导IT行业继续发展的动力则来自于数据。   
1　数据的重要性   
2　数据的统计和信息技术   
3　为什么需要大数据   
4　小结   
附录　计算复杂度　295   
第二版后记　299   
索引　302 

[查看全部↓](https://item.jd.com/11572052.html#detail-tag-id-6)

**精彩书摘**

　　《数学之美（第二版）》:  
　　第二，页面的分析和URL的提取。   
　　在上一节中提到，当一个网页下载完成后，需要从这个网页中提取其中的URL，把它们加入到下载的队列中。这个工作在互联网的早期不难，因为那时的网页都是直接用HTML语言书写的。那些URL都以文本的形式放在网页中，前后都有明显的标识，很容易提取出来。但是现在很多URL的提取就不那么直接了，因为很多网页如今是用一些脚本语言（比如JavaScript）生成的。打开网页的源代码，URL不是直接可见的文本，而是运行这一段脚本后才能得到的结果。因此，网络爬虫的页面分析就变得复杂很多，它要模拟浏览器运行一个网页，才能得到里面隐含的URL。有些网页的脚本写得非常不规范，以至于解析起来非常困难。可是，这些网页还是可以在浏览器中打开，说明浏览器可以解析。因此，需要做浏览器内核的工程师来写网络爬虫中的解析程序，可惜出色的浏览器内核工程师在全世界数量并不多。因此，若你发现一些网页明明存在，但搜索引擎就是没有收录，一个可能的原因是网络爬虫中的解析程序没能成功解析网页中不规范的脚本程序。   
　　第三，记录哪些网页已经下载过的小本本— URL表。  
　　……