

<<数学之美>>

图书基本信息

书名：<<数学之美>>

13位ISBN编号：9787115282828

10位ISBN编号：711528282X

出版时间：2012-5

出版时间：人民邮电出版社

作者：吴军

页数：272

字数：248000

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>

## &lt;&lt;数学之美&gt;&gt;

## 前言

去年我曾经给吴军的《浪潮之巅》写序，今年很高兴得知他的《数学之美》也即将出版了！和《浪潮之巅》一样，《数学之美》也是当年作为Google资深研究员的吴军在谷歌黑板报上应邀撰写的一系列文章。

说实在的，刚开始，黑板报的版主还有点担心这个系列会不会让读者觉得太理论而感到枯燥，但很快这个顾虑就被打消了。

《数学之美》用生动形象的语言，结合数学发展的历史和实际的案例，谈古论今，系统地阐述了与现代科技领域相关的重要的数学理论的起源、发展及其作用，深入浅出，受到广大读者尤其是科技类人士的喜爱。

之前就曾说过，在我认识的顶尖研究员和工程师里，吴军是极少数具有强大叙事能力和对科技、信息领域的发展变化有很深的纵向洞察力，并能有效归纳总结的人之一。

在《数学之美》里，吴军再次展示了这一特点。

与《浪潮之巅》不同的是，这次吴军集中阐述了他对数学和信息处理这些专业学科的理解，尤其是在语音识别、自然语言处理和信息搜索领域多年来的积累。

从数字和信息的由来，到搜索引擎对信息处理的背后的数学原理，到与搜索相关的众多领域后面的奇妙的数学应用，吴军都娓娓道来。

他把数学后面的本质思维写得透彻、生动。

不得不说，他的文字，引人入胜，也确实让我们体会到数学的美。

在他的笔下，数学不是我们一般联想到的枯燥深奥的符号，而是实实在在源于生活的有趣的现象和延伸。

数学，其实无处不在，而且有一种让人惊叹的韵律和美！

伽利略曾经说过，“数学是上帝描写自然的语言”；爱因斯坦也曾说过，“纯数学使我们能够发现概念和联系这些概念的规律，这些概念和规律给了我们理解自然现象的钥匙。

”我多年来一直也对信息处理、语音识别领域有着一定的研究，深深体会到数学在所有科学领域起到的基础和根本的作用。

“哪里有事，哪里就有美”。

在这里，我把《数学之美》真诚推荐给每一位对自然、科学、生活有兴趣有热情的朋友，不管你是搞理科还是搞文科的，读一读数学的东西，会让你非常受益，同时能感受到宇宙和世界的美好与奇妙。

吴军把之前谷歌黑板报上的“数学之美”系列文章编辑成现在的这本书，花费了大量的心血和时间。

他本着十分严谨的态度，在繁忙的工作之余，补充了之前的系列，并几乎重写了所有的文章，既照顾了普通读者的兴趣，又兼顾了专业读者对深度的要求，很让人钦佩。

有时我在想，现在的社会多了一点压力和浮躁，少了一点踏实和对自然科学本质的好奇求知。

吴军的这本《数学之美》真的非常好。

非常希望吴军今后能写出更多这样深入浅出的好书，它们会是给这个社会 and 年轻人最好的礼物。

李开复 2012年 于创新工场

## <<数学之美>>

### 内容概要

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，并被热情的读者广为传播，得到高度评价。

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，才真正明白“数学是科学的皇后”这句名言。

今年，作者吴军博士几乎把所有的文章都重写了一遍，为的是能把高深的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。

经过改写和重构后，《数学之美》在整体和细节的度上控制得更好。

希望读者通过具体的例子学到的是思考问题的方式，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。

同时书中也留了很多问题给愿意钻研的人做进一步深入思考。

## 作者简介

吴军博士，前腾讯公司主管搜索、在线广告和云计算基础架构的副总裁，毕业于清华大学（本科、硕士）和美国约翰·霍普金斯大学（博士）。

在清华大学和约翰·霍普金斯大学期间，吴军博士致力于语音识别、自然语言处理，特别是统计语言模型的研究。

他曾获得1995年全国人机语音智能接口会议的最佳论文奖和 2000年Eurospeech的最佳论文奖。

吴军博士于2002年加入Google公司。

在Google，他和Amit

Singhal（美国工程院院士，世界著名搜索专家）、Matt

Cutts（Google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Google工程奖。

2003年，他和Google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Google中日韩文搜索算法的主要设计者。

在Google其间，他还领导了许多研发项目，包括许多与中文相关的产品和自然语言处理的项目，并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。

吴军博士在国内外发表过数十篇论文，并获得和申请了十余项美国和国际专利。

他的《浪潮之巅》一书深受业界的好评。

他于2007年起，担任风险投资基金中国世纪基金的董事。

2011年起当选为约翰·霍普金斯大学工学院董事会董事，并在该校的国际事务委员会担任顾问。

他是国家重大专项“新一代搜索引擎和浏览器”项目的总负责人，从2012年起担任工信部的专家和顾问。

## &lt;&lt;数学之美&gt;&gt;

## 书籍目录

i 出版说明

v 序言1

ix 序言2

xi 前言

### 第1章 文字和语言 vs 数字和信息

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

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?? 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

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 谈谈中文分词

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：工程上的细节问题

3 小结

### 第5章 隐含马尔可夫模型

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。

## &lt;&lt;数学之美&gt;&gt;

- 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

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 延伸阅读：信息论在信息处理中的应用
- 4 小结

## 第7章 贾里尼克和现代语言处理

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的索引

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

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的民主表决式网名

网页排名技术是早期的杀手锏，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

## &lt;&lt;数学之美&gt;&gt;

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

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。

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

页排名技术

2 延伸阅读：TF-IDF 的信息论依据

3 小结

## 第12章 地图和本地搜索的最基本技术 — 有限制

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·士

在所有轻武器中最有名的是冲锋枪，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。

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

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

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

## &lt;&lt;数学之美&gt;&gt;

## 3 小结

## 第16章 信息指纹及其应用

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

原理

密码学的根本是信息论和数学。

没有信息论指导的密码

是很容易被破解的。

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

题

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。

## 第19章 谈谈数学模型的重要性

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经常常是曲折的。

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——谈谈型

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反作弊问题着广泛的应用。

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 最大熵模型的训练



## &lt;&lt;数学之美&gt;&gt;

## 3 小结

## 第21章 拼音输入法的数学原理

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

## 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们

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

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

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 条件随机场和句法分析

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

## &lt;&lt;数学之美&gt;&gt;

么联系呢？

- 1 句法分析计算机算法的演变
- 2 条件随机场
- 3 小结

## 第26章 维特比和他的维特比算法

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，因为如今基于的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 再谈文本自动分类问题 — 期望最大化

算法

只要有一些训练数据，再定义一个最大化函数，采用算法，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

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

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

<<数学之美>>

3 小结  
附录  
后记  
索引

## &lt;&lt;数学之美&gt;&gt;

## 章节摘录

世间万般表象的背后皆有规律。

科学研究的真正目的，在于找到这些规律并能够举一反三地应用这些规律。

对于自然科学和工程学，包括计算机科学和通信，数学是描述这些规律性最好的工具。

做事情的方法有对与错、好和坏之分，从长期看只有采用对的方法、好的方法才能做出好的产品和服务。

而任何一个好的方法背后都有充分的理论依据，都有正确的方法论。

本书的一个目的在于揭示信息和自然语言处理以及它们在通信和互联网各种应用（语音识别、机器翻译、搜索、分类等）中的数学原理，从而帮助读者认识到：解决这些问题，好的方法是什么。

毫无疑问，解决这些问题需要智能，而计算机本身并没有智能，它有的只是巨大的计算能力。

只有找到描述这些问题的数学模型，才能利用计算机解决这些需要智能的问题，这样，计算机看上去也似乎聪明起来了。

数学之美，首先在于其内容或许复杂而深奥，但形式常常很简单。

简单=美，这不仅在科学和工程上成立，而且在产品开发和设计上也是如此，这和设计大师乔布斯“少即是多”的原则不谋而合。

同时，数学之美还在于数学原理的通用性和普遍性。

一个好的数学模型，常常能解决一系列，甚至是许多看似毫不相干领域的实际问题。

在数学上的一点突破，可以带动很多领域和行业的进步。

数学给人的印象常常是抽象而深奥、难以理解和应用不明确的。

产生这个这个误解的原因是缺乏从数学到应用的桥梁，本书希望抛砖引玉，期待着能有更多的人一起来搭建这座桥梁。

## &lt;&lt;数学之美&gt;&gt;

## 媒体关注与评论

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》创始人 最初看到《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 第一次接触吴军老师的“数学之美”系列，是在搜索bloom filter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，最重要的是理解知识的由来；其次，任何复杂的问题最终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最好的文章之一，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

## <<数学之美>>

### 编辑推荐

《浪潮之巅》作者、腾讯公司副总裁吴军最新力作《数学之美》由创新工场董事长兼首席执行官李开复倾力作序推荐。

《数学之美》的创作源自点击超百万的谷歌黑板报专题博客，吴军老师应出版要求重新编写。

在《数学之美》中吴军老师带领你领略数学之美，架起从数学到应用的桥梁，教会你如何化繁为简，如何用数学解决工程问题，如何打破思维定式不断思考创新。

《数学之美》是献给每一位对自然、科学、生活有兴趣的朋友的礼物。

本书荣获第八届文津图书奖。

## &lt;&lt;数学之美&gt;&gt;

## 名人推荐

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》创始人最初看到《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》作者第一次接触吴军老师的“数学之美”系列，是在搜索bloom filter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，最重要的是理解知识的由来；其次，任何复杂的问题最终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最好的文章之一，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请看看本书的章节名称，有“文字和语言vs数字和信息”、“自然语言处理——从规则到统计”、“统计语言模型”、“谈谈中文分词”、“隐含马尔可夫模”、“贾里尼克和现代语言处理”、“布尔代数和搜索引擎”、“图论和网络爬虫”、“余弦定理和新闻的分类”、“信息指纹及其应用”、“逻辑回归和搜索广告”等，似乎太过专业性，实际上具有高中和大学低年级的同学们都能看得懂，当然本书因此也可以称得上是“高级科普”。

总之，本书是在介绍信息技术背后的数学原理，以及这些技术和原理背后的科学家的故事和他们从事科学研究的方式方法。

故事生动翔实而富有教益，是一部优秀的科学方法论的启蒙读物。

——王渝生 文津图书奖科普类图书专家评审委员会主任，中国科技馆原馆长，北京市科协原副主席人民邮电出版社《数学之美》，作者吴军是非数学专业的理工科出身的企业界人士，他不落窠臼，独辟蹊径，站在信息时代的高度，领略数学应用于工程技术中的实质内容和内在美感，具有科学方法论的启蒙意义。

——王渝生 文津图书奖科普类图书专家评审委员会主任，中国科技馆原馆长，北京市科协原副主席

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>