真正统治世界的十大算法

算法爱好者 2016-05-25

> 英文：Marcos Otero
> 译者：伯乐在线 - Justin Wu
> 链接：http://blog.jobbole.com/70639/

不久前的某一天，我在浏览Reddit发现了一篇有趣的文章《统治世界的十大算法》，作者George Dvorsky在那篇文章中试图解释算法之于当今世界的重要性，以及哪些算法对人类文明最为重要。

此时此刻，如果你已经学过算法的话，那么在你阅读那篇文章时，你脑海中所浮现的第一件事也许是“作者是否明白算法是什么？”或是“**Facebook**的新闻提要是一种算法？”，因为如果Facebook的新闻提要也算是一种算法的话，那么最终你可以把几乎所有的东西都归类为算法。因此，在本文中我会试着去解释什么是算法，以及哪十个（也许更多）算法是真正统治世界的。

什么是算法？

> 直白地说，算法就是任何明确定义的计算过程，它接收一些值或集合作为输入，并产生一些值或集合作为输出。这样，算法就是将输入转换为输出的一系列计算过程。来源：**Thomas H. Cormen, Chales E. Leiserson (2009),** 《算法导论第三版》。**

简而言之，我们可以说算法就是用来解决一个特定任务的一系列步骤（是的，不止计算机在使用算法，人类也同样如此）。目前，一个有效的算法应该含有三个重要特性：

1. 它必须是有限的：如果你设计的算法永无休止地尝试解决问题，那么它是无用的。

2. 它必须具备明确定义的指令：算法的每一步都必须准确定义，在任何场景下指

令都应当没有歧义。

3. 它必须是有效的：一个算法被设计用以解决某个问题，那么它就应当能解决这个问题，并且仅仅使用纸和笔就能证明该算法是收敛的。

还有一个要点需要指出，算法不仅仅在计算机科学中使用，同时也存在于数学领域中。事实上，首个被记载的数学算法要追溯到公元前1600年，古巴比伦人开发了已知最早的算法，用作因式分解和计算平方根。这里，我们回答了前面所提到的那篇文章中的第一个问题，它认为算法是计算机范畴的实体，但如果你知晓算法这个词的真正内涵的话，真正统治世界的十大算法也能在数学书籍中找到（加法、减法、乘积等等）。

不过在这篇文章中，让我们将算法的定义限定在计算机算法上，所以剩下的问题是：哪十个算法统治了世界？在此我整理了一个小型列表，排名不分先后。

1. 归并排序，快速排序和堆排序

Name	Best	Average	Worst	Memory	Stable
Bubble sort	n	n^2	n^2	1	Yes
Selection sort	n^2	n^2	n^2	1	No
Insertion sort	n	n^2	n^2	1	Yes
Merge sort	$n \log n$	$n \log n$	$n \log n$	worst case is n	Yes
In-place merge sort	—	—	$n (\log n)^2$	1	Yes
Quicksort	$n \log n$	$n \log n$	n^2	$\log n$ on average, worst case is n	typical in-place sort is not stable;
Heapsort	$n \log n$	$n \log n$	$n \log n$	1	No

哪个排序算法最好？这取决于你的需求，这也是为什么我要将这三个使用频率较高的排序算法置于一处的原因。可能你比较偏爱其中一个，但它们都是同等重要的。

归并排序算法是目前为止我们拥有的最重要的算法之一。它是一种基于比较的排序算法，使用分治法解决那些原本复杂度为O(N^2)的问题。归并排序是由数学家

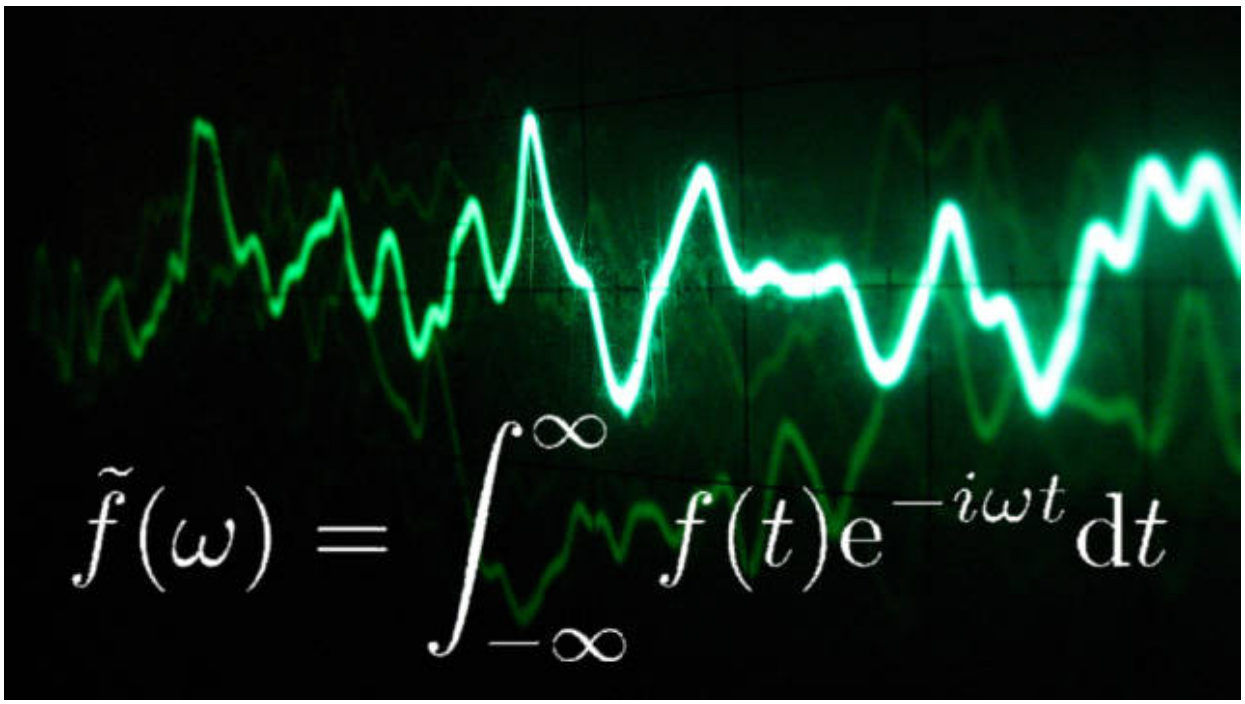
John von Neumann于1945年发明的。

快速排序是解决排序问题的另一种途径，它使用就地分解算法，同时它也是一种分治算法。这个算法的问题在于它是不稳定的排序算法，但它在基于内存的数组排序上确实非常高效。

最后，堆排序算法使用一个优先队列降低数据的查找时间，它也是一种就地排序算法，同样也是不稳定的排序算法。

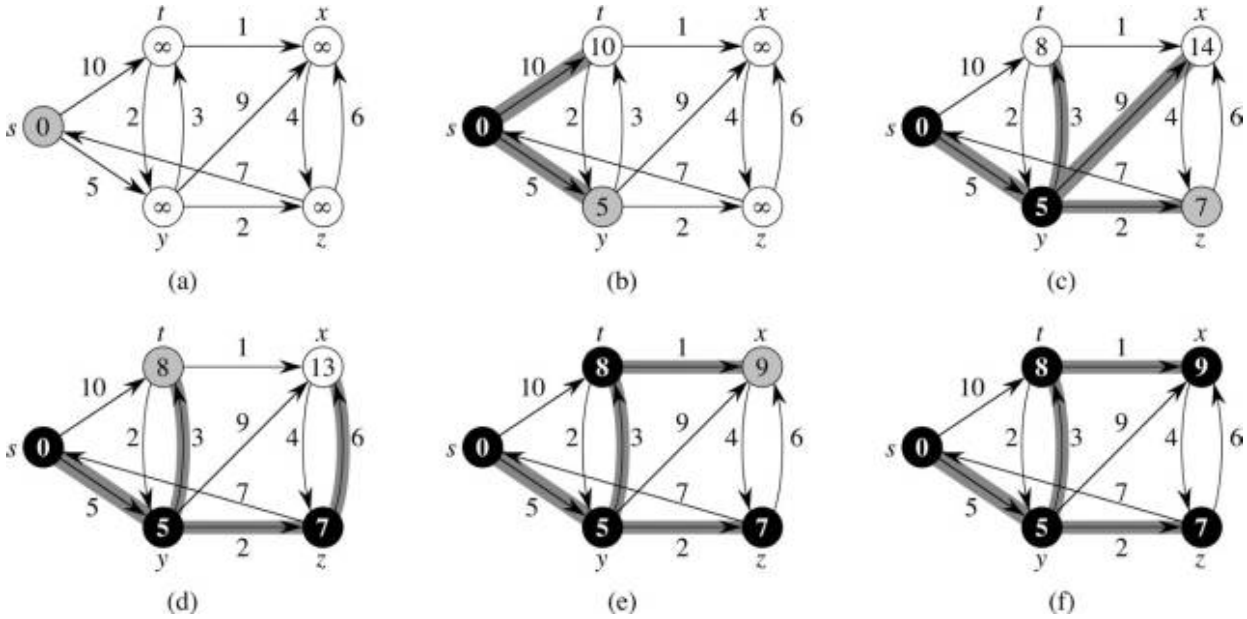
相较于曾经使用的其他排序算法（如冒泡排序），上述算法带来了显著的改进。事实上，多亏了它们，今天我们才有了数据挖掘、人工智能、链接分析，以及世界上大部分的计算机工具，也包括网络在内。

2. 傅立叶变换与快速傅立叶变换



整个数字世界都在使用这些简单而又强大的算法，将信号从频域转换为时域，反之亦然。事实上，正是归功于这些算法，你才能看到这篇文章。互联网、你的WIFI、智能手机、电话、计算机、路由器、卫星，几乎所有内置计算机的东西都会以各种方式使用这些算法实现各自的功能。如果你没有学习这些重要的算法，你将无法获得电子、计算机或通信方面的学位。

3. Dijkstra 算法



毫无不夸张地说，如果没有这个算法，当今互联网将无法有效工作。这是一种图搜索算法，它被广泛应用在能够建模为图的问题中，用以找出两个节点之间的最短路径。

目前，即便我们已经拥有了解决最短路径问题的更好方法，Dijkstra 算法依然在那些重视稳定性的系统中得到应用。

4. RSA算法

如果没有信息加密和网络安全，互联网不会像现在那么重要。你可以认为“安全问题理所当然应该是美国国家安全局和其他情报机构的事情”或“你认为你身处在互联网是安全的，这太天真了”。但是，人们需要在他们花钱时保有安全感，毕竟你不会在网络服务器上输入你的信用卡号，如果你知道它是不安全的话。

在信息加密领域，有一个算法始终是世界上最重要的算法之一，它就是RSA算法。这个算法是由RSA公司的创始人所建立的，它使信息加密惠及千家万户，奠定了当今信息加密的运作基础。RSA算法用来解决一个简单而又复杂的问题：怎样在不同平台和终端用户之间共享公钥，继而实现信息加密（我想说明一下这个问题还没完全解决，我想我们需要基于这个方向做更多工作）。

5. 安全哈希算法

准确地说，它不能称之为是算法，它是美国国家标准暨技术学会定义的加密散列函数族中的一员，但是这族算法对整个世界的运作至关重要。从你的应用商店，你的邮件，你的杀毒软件，到你的浏览器等等，所有这些都在使用安全哈希算法，它能判断你是否下载了你想要的东西，也能判断你是否是中间人攻击或网络钓鱼攻击的受害者。

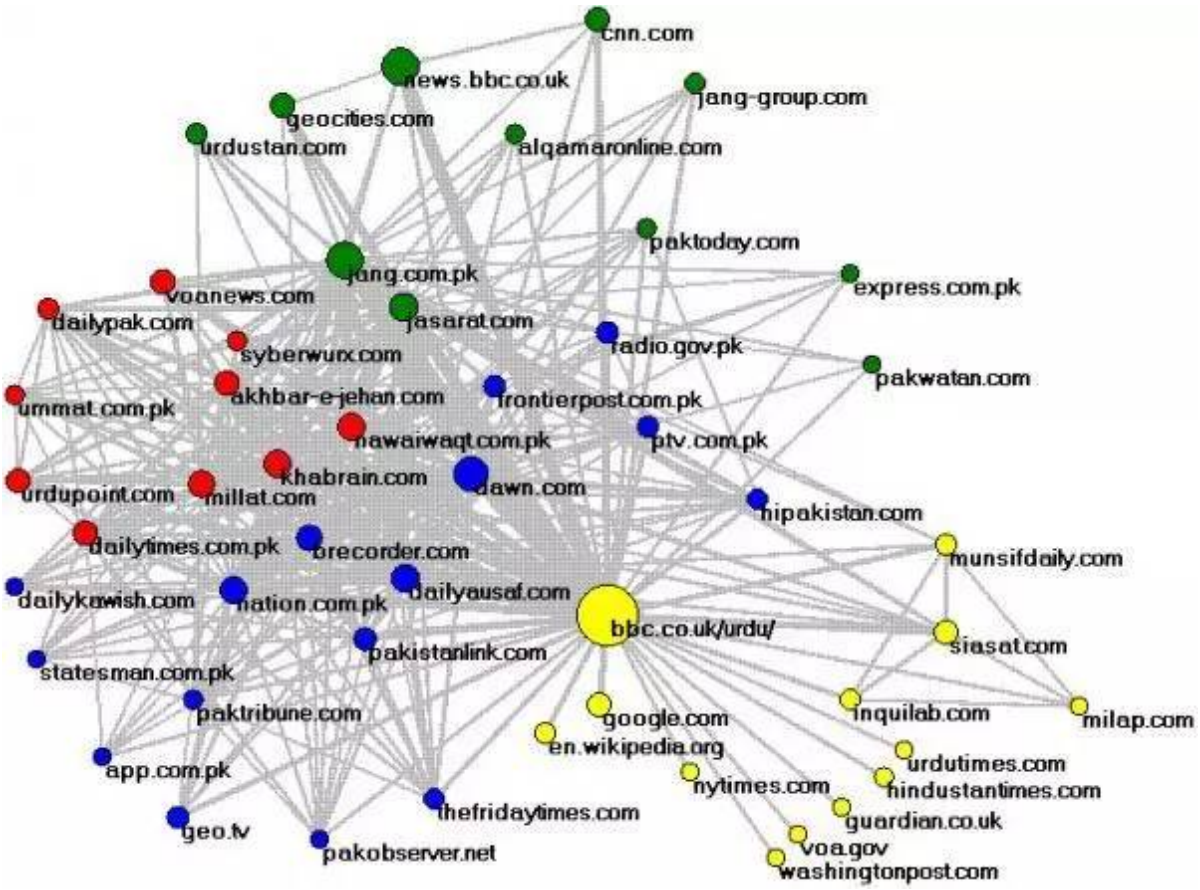
6. 整数因式分解

这是在计算机领域被大量使用的数学算法，没有这个算法，信息加密会更不安全。该算法定义了一系列步骤，得到将一合数分解为更小因子的质数分解式。这被认为是一种FNP问题，它是NP分类问题的延伸，极其难以解决。

许多加密协议（如RSA算法）都基于这样一个原理：对大的合数作因式分解是非常困难的。如果一个算法能够快速地对任意整数进行因式分解，RSA的公钥加密体系就会失去其安全性。

量子计算的诞生使我们能够更容易地解决这类问题，同时它也打开了一个全新的领域，使得我们能够利用量子世界中的特性来保证系统安全。

7. 链接分析



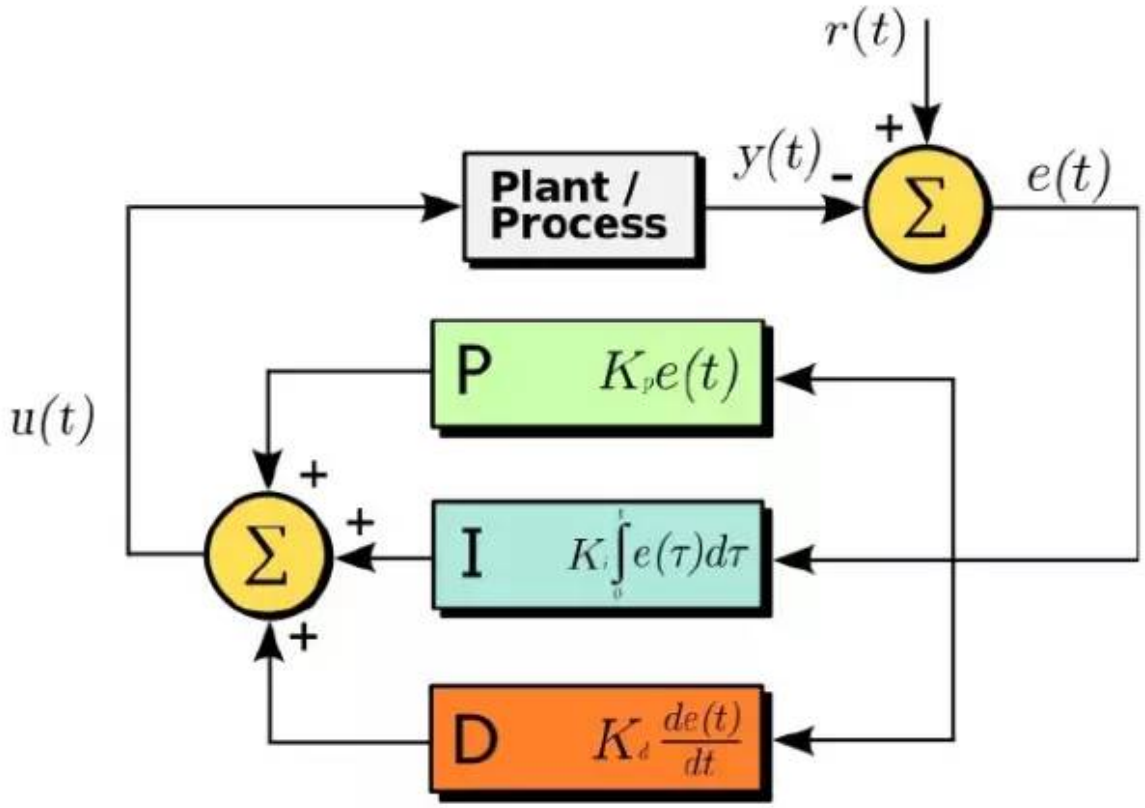
在互联网时代，分析不同实体间的关系是相当重要的。从搜索引擎，社交网络，到营销分析工具，每个人都在不停寻找互联网的真正结构。
有证据显示，链接分析是公众心目中伴随着最多谬见和误解的算法之一。这里的问题在于，有很多不同的方式可以进行链接分析，也存在很多特性使这些算法看起来有细微的区别（这些区别允许该算法独立申请专利），但它们本质上是类似的。

链接分析背后的理念非常简单，以矩阵形式描绘出一张图，将问题转换为特征值问题。特征值是一种很好的渠道，它有助于展现图的结构以及每个节点的相对重要性。该算法是由Gabriel Pinski和Francis Narin于1976年建立的。

谁在使用这个算法？Google的Page Rank算法，Facebook向你展示的新闻提要（这就是为什么Facebook的新闻提要不是算法，只是使用算法的结果而已），Google+和Facebook的好友推荐，LinkedIn的工作和联系人推荐，Netflix和Hulu的电影，YouTuBe的视频，等等。虽然每个都有不同的目标和参数，但它们背后的数学理念是相同的。

最后，我想说明一点，尽管看上去Google是第一家使用这类算法的公司，然而在1996年（Google之前两年），Robin Li（李彦宏）所建立的一个小型搜索引擎“RankDex”就已经在它的网页排名机制中使用了这项理念。后来，HyperSearch的创始人Massimo Marchiori基于各网页之间的关系使用了另一种网页排名算法。（Google在它的专利中提到了这两位创始者）

8. 比例积分微分算法



你是否曾经用过飞机、汽车、卫星服务或手机网络？你是否曾经在工厂工作或是看见过机器人？如果回答是肯定的，那么你应该已经见识过这个算法了。

大体上，这个算法使用一种控制回路反馈机制，将期望输出信号和实际输出信号之间的错误最小化。无论何处，只要你需要进行信号处理，或者你需要一套电子系统，用来自动化控制机械、液压或热力系统，这个算法都会有用武之地。

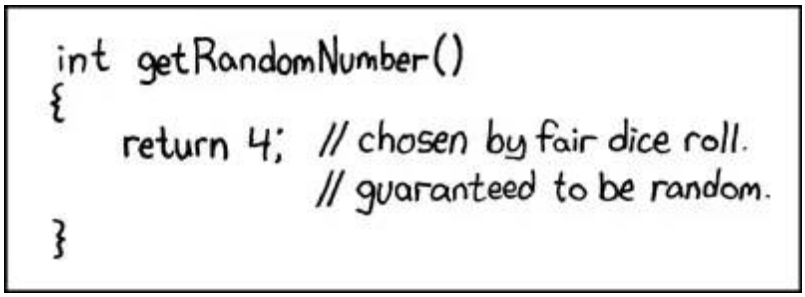
可以这样说，如果没有这个算法，现代文明将不复存在。

9. 数据压缩算法

要判断哪种数据压缩算法最为重要是很困难的，因为它取决于不同的应用环境。它们可以应用在zip和mp3上，也可以应用在JPEG和MPEG-2上。但众所周知，在所有结构中这些算法都极其重要。

除了显而易见的zip文件，在哪我们能够找到这些算法？这张网页就进行了数据压缩并被下载到你本地，同时我们还能在电子游戏、视频、音乐、数据存储、云计算、数据库等等地方找到这些算法。可以说，数据压缩算法处处可见，它们使系统成本更低、效率更高。

10. 随机数生成



现在我们还没有一个“真正的”随机数生成器，但我们已经有了一些伪随机数生成器，这够用了。随机数生成器的用途非常广泛，从互联联络、数据加密、安全哈希算法、电子游戏、人工智能、优化分析，到问题的初始条件、金融等等，都有它们的身影。

最后，我想强调一下，上面这个列表经供参考，它并不完整。因为在机器学习、矩阵乘法、分类化等领域也有一些算法，它们对我们的世界同样重要，但在这里还没有提到。

算法爱好者

专注算法相关内容

微信号：AlgorithmFans

长按识别二维码关注

伯乐在线 旗下微信公众号

商务合作QQ：2302462408