

# Python 正在慢慢褪色 大数据文摘出品

编译 陈之炎 coolboy

自从 20 世纪 90 年代初发布以来，Python 一直相当火爆，在这二十多年里，它的流行程度远远超过了 C、C#、Java 甚至 Javascript。

虽然 Python 在数据科学和机器学习领域占主导地位，甚至是科学和数学计算领域的主角，但与 Julia、Swift 和 Java 等语言相比，它确实有些欠缺。

## 为什么 Python 如此受欢迎？

Python 迅猛发展背后的一个主要驱动力是它学习起来相当容易，使用起来功能强大。对于初学者来说，像 C/C++ 这样有难度的编程语言，不熟悉语法的人都会敬而远之，所以 Python 显得非常有吸引力。

语言的核心是代码可读性，Python 的语法简洁而富有表现力，开发人员无需编写大量代码，便可以表达想法和概

念（C 或 Java 等底层语言的情况也是如此）。它非常简单，并可以与其他编程语言无缝集成（比如将 CPU 密集型任务交付予 C/C++）。用 Python 语言开发会给开发人员带来许多好处。

Python 应用广泛的另一个原因是它被企业大量使用（包括 FAANG）。今天，对于你能想象到的任何项目，都可以找到一个相应的 Python 包——科学计算的有 Numpy，机器学习的有 Sklearn，计算机视觉的有 Caer。

Python 的弱点它很慢，太慢了。不用多说，速度通常是开发人员最为关注的焦点之一。不仅是现在，将来的很长一段时间都会如此。

Python “缓慢”的主要原因，可以归结为以下两点——首先，Python 是解释性的语言，而不是编译型的语言，这终究会导致执行时间的缓慢；其次，它是一种动态的语言（变量的数据类型在执行期间由 Python 自动推断）。

初学者经常对“Python 的确太慢”的说法产生争议。事实上，的确如此，但也并非完全这样。

例如，Python 的机器学习库 TensorFlow 实际上是用 C++ 编写的，在 Python 中进行调用而已，是在 C++ 的上层实现的一个 Python “封装器”。Numpy 是如此，Caer 也是类似的形式。

## Python 有一个 GIL(L)

Python 缓慢的主要原因之一是 GIL(Global Interpreter Lock) 的存在——它一次只允许一个线程执行。这虽然有助于提高单线程的性能，但限制了并行性，而开发人员为了提高速度必须实现多线程处理程序

不是内存密集型任务的最佳选择 当对象超出范围时，Python 会自动进行垃圾收集。其目的是消除 C 和 C++ 在内存管理中涉及的许多复杂性。由于指定的数据类型缺乏灵活性，Python 消

耗的内存量可能会迅速暴增。

此外，在运行时，一些 BUG 可能会被 Python 忽略，最终成为延缓开发过程的主要诱因。

## 在移动计算中表现牵强

随着从台式机到智能手机的迁移，显然，人们需要为手机软件构建更强大的语言。虽然 Python 在计算机和服务器平台上有不错的表现，但由于缺乏强大的移动计算处理能力，在移动开发中它往往会失去优势。

近年来，（Python）在这方面有了很大的进步，但这些新增的库与他们强大的竞争对手相比，如 Kotlin、Swift 和 Java，还相去甚远。

其他语言的兴起 最近，像 Julia、Rust 和 Swift 这些新出现的语言，借用了 Python、C/C 和 Java 的许多良好设计概念——Rust 确保了运行时的内存安全和并行，并提供了

与 Web Assembly 的一样的互操作性；由于它支持 LLVM 编译器工具链，Swift 几乎和 C 一样快；Julia 提供了用于 I/O（输入/输出）密集任务的异步 I/O，而且速度惊人。

## 结论

Python 并非最好的编程语言，它无法取代 C/C++ 和 Java。它被打造成一种通用的编程语言，强调可读性、以英语为中心的语法。人们利用它可以快速开发程序和应用程序。

和其他语言一样，Python 只是一种工具。某些应用场景下，它是最好的工具；在另一种场景下，它或许不是。在大多数情况下，它“挺好用的”。

那么，作为一种编程语言，Python 会消亡吗？我不这么认为。

它逐渐失去魅力了吗？啊，也许只是一点点，只是一点点。👍

（来源：medium）

# SpaceX 载人飞船再升空 首次搭载四位宇航员，没受马斯克新冠测试影响

北京时间 11 月 16 日 8 时 27 分，SpaceX 龙飞船又一次成功升空了，这一次载着四位宇航员。

四位宇航员三位来自美国，一位来自日本，顺利对接后，他们将在空间站进行 6 个月的太空探索。

此次载人龙飞船的发射，也标志着 SpaceX 与 NASA 的商业载人航天计划正式开启。

美国东部时间 11 月 16 日晚 7 点 27 分，名为“坚韧”（Resilience）的载人龙号太空舱发射升空，开始了与国际空间站对接的大约 27 个小时的旅程。

这次要顺利到达空间站，SpaceX 的龙飞船需要经过六个步骤：升空、入轨、点火、启动、靠近、对接和加压。

今年 5 月，搭载两名宇航员的“龙”号太空舱成功进行了试飞，没有出现任何重大问题，成为美国自 2011 年以来首次发射载人航天飞船，也是美国进入由商业主导太空探索的历史性时刻。

5 月份的成功发射也为本次的试飞扫清了道路。之后的半年时间，SpaceX 也更新了航天器，升级了太阳能阵列，燃料泵和隔热罩。

NASA 上周正式决定，使用 SpaceX 的猎鹰 9 号火箭和龙太空舱安全运送宇航员，用于为国际实验室服务的常规船员轮换和其他操作飞行。

## 美国首次送出 4 位宇航员， 搭载首位黑人宇航员

此次乘坐龙飞船登上太空的 4 位宇航员分别是：迈克尔·霍普金斯 (Michael Hopkins)、维克托·格洛弗 (Victor Glover) 和 香农·沃克 (Shannon Walker)，以及野口聪一 (Soichi Noguchi)。他们将在太空执行长达六个月的任務。

这次任务也是美国宇航局首次将四名宇航员送入太空舱。升空成功后，在空间站为期六个月工作期间，几位宇航员将进行一系列的各种科学实验的工作，包括太空制造和健康研究。野口聪一此前曾乘坐过航天飞船。

值得一提的是，四名宇航员中还包含一位女性宇航员 Shannon Walker，她于 2004 年被美国国家航空航天局 (NASA) 选为宇航员。她拥有莱斯大学的物理学学士学位，科学硕士学位和空间物理学博士学位。

此次任务还包括维克托·格洛弗 (Victor Glover)，他于 2013 年被选为宇航员，将成为首位作为宇航员在空间站长时间停留的黑人。执行这次任务的四名宇航员中有三名此前曾进入过太空，但对宇航员维克托来说，这是他第一次进入太空。

## SpaceX 联手 NASA， 美国宇航时代复兴？

SpaceX 的载人航天计划是在 NASA

的 Commercial Crew 计划下进行的，NASA 选择了两家私营公司来建造宇航员发射系统，将宇航员从美国本土运送到国际空间站。2014 年，SpaceX 和波音公司一同被 NASA 选中，来创建各自的系统。

从 2014 年到现在，SpaceX 为了完成一系列里程碑式的任务一直都在努力着，包括全自动的无人国际空间站交会任务，以及一个用于保护宇航员的发射台中止和发射后中止的紧急安全系统。在去年 3 月 Demo-1 任务期间，尽管所有实际的发射，对接和着陆都由 SpaceX 的完全自主的软件和导航进行处理，但宇航员还短暂地接管了手动控制。

而在今年 5 月，SpaceX 的龙飞船和猎鹰 9 号火箭成功搭载两名宇航员登上太空后，SpaceX 成为了美国历史上首个实现商业航天的私营公司。

自 2011 年航天飞机退役以来，美国国家航空和宇航局 (NASA) 设想最终依靠私人开发和操作的太空舱来执行人类轨道任务。与此同时，NASA 在等待国内方案时完全依赖俄罗斯的火箭和太空舱。

太空探索技术公司的发射费用预计将比克里姆林宫的费用少 40% 左右。美国更广泛的目标是释放机构资金，以便更深入地探索太阳系。

而今天这次载人任务是自 2011 年航天飞机项目结束以来，NASA 首次在常规操作任务中向太空发射宇航员。这标志着美国载人航天能力的复兴，并有望成为 SpaceX 和龙飞船所承担的多项载人飞行任务中的第一个——包括 NASA 的飞行任务和商业客户组织的飞行任务。

当 NASA 确定 SpaceX 已经为执行任务做好了准备时，该机构人类探索项目负责人凯西·吕德斯 (Kathy Lueders) 说：“你们已经向我们证明，能够提供满足我们要求的载人运输能力。”与上世纪 70 年代开发的航天飞机相比，NASA 显著提高了龙飞船太空舱的安全标准。

上周，马斯克称 NASA 的认证“是一项巨大的荣誉，鼓舞了我们重返月球、前往火星、最终帮助人类成为多行星的努力的信心”。

尽管当天的大部分时间里发射台附近都持续有云和阵雨，但最终天气还是足够好，适合发射，在紧急着陆地点也是可以接受的。倒计时在发射前的准备工作中只出现了一个小小的技术故障，随后猎鹰 9 号火箭和泪滴形状的宇宙飞船将以 208 英尺（约合 12 米）的高度升空。

到目前为止，SpaceX 已经从联邦政府获得了超过 31 亿美元的商用乘员计

划资金，其竞争对手波音公司 (Boeing Co) 估计也获得了 48 亿美元资金。

波音公司在 2019 年底遭遇了一次失败的飞行测试，当时一个软件问题导致其搭载无人的 Starliner 太空舱无法到达正确的轨道，与空间站对接。波音公司预计将在未来几个月进行另一次试飞，也不会载人，宇航员可能会在 2021 年的某个时候在空间站执行示范任务。

## 发射前两天马斯克被测试出新阳性， 发射状态曾被推迟

除了天气问题，这次发射还受到疫情影响。马斯克在 11 月 12 日发推特称，有可能感染新冠。NASA 最初表示，由于马斯克的健康状况，此次发射的状态尚不确定，但后来由于与之无关的天气问题，将发射从周六推迟到周日。

马斯克在 Twitter 上发消息称：“进行了四次新冠检测，两阴两阳。”但他没有说明他是否收到了更多的测试结果。“我的症状是轻微感冒。”

之后，NASA 发布规定要求马斯克必须自我隔离。但据布里斯登廷称，由于马斯克没有接触到任务所需的任何人员，发射依然照常进行。

其实 10 月下旬以来，宇航员一直被隔离，以保护他们免受病毒影响。

SpaceX 的发言人没有回复记者就马斯克的状况置评的请求。👍

# 人人都说“数据直觉”很重要，但它到底是个啥？又该如何培养？

编译 张大笔笔

几天前，一位同事让我解释下“数据直觉”的含义，我突然发现虽然自己一直在强调数据直觉，却从来没有真正的定义过它。

数据直觉能力是我面试新数据科学家所关注的三项技能之一（其他两项是统计和技术能力）。其实我只是在 2020 年前 9 个月建立了 Mozilla 的数据直觉概念，现在却突然惊讶地发现无法为自己要培养的东西做出解释。

所以现在来补充一下！以下为数据直觉的定义：数据直觉是对误导性的数据和分析的敏感度。

换句话说，如果某人具有很强的数

据直觉，他就很难被数据误导，你也可以认为是对黑暗数据艺术的防御能力。

那具体在实践中表现在哪些方面呢？

## 数据嗅觉

具有强烈数据直觉的人可以迅速闻到“数据味道”（与“代码气味”相近）。这种数据问题不一定会影响分析，但结果肯定会令人质疑。例如：

分析结果显示一项看似随意的指标 4 天内的保留率提高了 0.5%。4 天保留时间从何而来？我们通常不是跟踪 7 天的保留时间吗？这是值得注意的一点。

分析报告的结果：此功能将保留率提高了 10%。但是，过去不是一直在努力将保留率提高 0.9% 吗？而且保留率已经不是 90% 吗？那么要如何获得并增加 10%？

这些都是极端的例子。通常情况下，问题只会更加微妙，有数据直觉的人会觉得结果怪怪的（这就是为什么将其称为“直觉”）。

显然，数据直觉与产品直觉是有关的，尽管这俩属于不同的技能。产品的直觉可以使结果与实际相符，并且可以更轻松地识别分析中的特殊要求。要知道保留率提高 10% 是荒谬的，我们需要知道的是用户已经保留得很好了（90% 的保留率）。

## 方法问题

强大的数据直觉还可以发现分析设计的问题。例如：作者是如何收集数据的？样本有代表性吗？是否需要实验来论证因果关系？

举个例子：一项分析报告表示，创建 Firefox 账户的用户比未创建 Firefox 的用户保留率高 10%。默认情况下，许多人将其解释为，如果我们花一些时间来帮助用户开设账户，则保留率会增加。有数据直觉的人们会认识到这个结果只是相关关系（并非因果关系）。

经常使用该产品的用户可能会停留更长的时间。开立账户的用户都是活跃用户，因此保留率更高。频繁使用 Firefox 的用户是更活跃的用户，保留率会更好。

我认为这种直觉不仅是很好地理解统计数据。强大的统计背景可以在阅读白皮书的方法部分时发现问题，可以让我对新闻头条中听到的结果有多信任，更可以帮助我确定结果是否足够真实。

## 不仅仅是怀疑主义

我几乎将数据直觉定义为一种怀疑态度，但这不是一个好的描述。怀疑主义过于强调结果了。直觉不仅仅是怀疑。它会将新数据纳入现有知识体系的一部分。大多数时候这意味着确定新的传入数据不一致，需要更多调查才能信任。其他时候则意味着需要比现有知识体系更具权威性的新数据来改变我们的观点。

## 你是怎么想的？

我想听听你的想法！之所以公开发布此定义，部分原因是我想用坎宁安定律，即获得正确答案的最佳方法是发布错误答案！这个数据直觉的定义能引起你的共鸣吗？👍

（来源：数据科学家 Ryan T. Harter 博客）