

数据科学的演变：从不为人知成为主流学科

作者：[刘永川 \(Alex Liu\)](#)，北京大学硕士和斯坦福大学博士，一直从事数据分析，曾任 IBM 分析服务的首席数据科学家。目前负责 RMDS 实验室，并担任哈佛数据科学评论等机构顾问。

数据科学：从不为人知到无处不在

今天，数据科学已成为我们生活中无处不在的一部分。无论是你每天使用的手机应用，还是你在社交媒体上看到的推荐内容，数据科学都在背后默默发挥着作用。然而，即便在短短的二十多年前，数据科学还只是一个默默无闻的小众领域。虽然早在 20 世纪 60 年代，“数据科学”这个词就已经在一些学术研究中出现了，但那时它并没有产生多大的实际影响。

起步阶段：从怀疑到探索

2004 年，我有幸先在一家技术咨询公司、后在一家产品搜寻公司 Shopzilla 担任数据科学家的角色，虽然短暂，但可能是当时屈指可数的数据科学家之一。那时，数据科学还是一个非常新颖的概念。我的合作者 Usama Fayyad，当时在雅虎成为首席数据科学家，据说是第一位使用首席数据科学家头衔的先驱者。然而，当时很多人对数据科学的存在感到疑惑，当时我们给别人数据科学家的名片时，得到第一个问题都是“什么是数据科学啊？”。他们认为，数据科学不过是统计学和计算机科学的混合体，没有理由将它单独列为一个学科。学术界和业界对这种跨学科尝试心存疑虑，担心它会模糊传统学科的界限。

但我们知道，随着数字时代的到来，数据量正在以惊人的速度增长，而传统的数据分析方法已经无法应对这些海量信息。企业逐渐意识到，必须有专门的人才来处理和解读这些数据。2013 年，IBM 发布的统计数据显示，全球 90% 的数据是在过去两年内产生的。这一令人震惊的数据增长速度，推动着数据科学从幕后走向了舞台中央。

转折点：数据科学进入大众视野

2012 年，我的另一位合作者、被一些媒体称为大数据之父的 Tom Davenport，与 D.J. Patil 在《哈佛商业评论》上发表了一篇文章，标题非常吸引人——《21 世纪最性感的工作：数据科学家》。这篇文章不仅强调了数据科学家的重要性，还成功地将这个职业推向了大众的视野。可以说，这篇文章帮助数据科学从小众走向了主流，成为了许多人梦寐以求的职业选择。

快速增长：从学术认可到政府重视

这篇文章发表后，数据科学的发展进入了快车道。2013 年，纽约大学开设了美国第一个正式的数据科学学位课程。这标志着学术界开始正式承认数据科学为一个独立学科。不久之后，全球各地的大学纷纷效仿，如今已有至少数百所高校开设了专门的数据科学学位课程。

2015 年，《21 世纪最性感的工作：数据科学家》的作者之一 D.J. Patil 被奥巴马政府任命为美国首位首席数据科学家，这进一步巩固了数据科学在国家层面的重要性。这一任命表明，数据科学不仅在商业领域举足轻重，在政府管理中也开始发挥关键作用。与此同时，我自己也在 2013 年作为数据科学家加入了 IBM，在职业生涯中不断发展，担任了 IBM 分析服务的首席数据科学家，见证并参与了企业对数据科学人才需求的急剧增长。

不仅如此，像纽约和洛杉矶这样的城市也开始聘请首席数据官，这反映出数据驱动治理的广泛趋势。这一变化表明，数据科学不仅仅是私营公司的工具，它也成为了公共部门高效运作的关键技术。

推动数据科学发展的三大引擎：数据、实际需求与计算能力

回顾数据科学的快速崛起，我认为这是由三股强大的力量推动的：越来越多的数据、企业对数据驱动解决方案的迫切需求，以及计算能力的飞速进步。

数据的爆炸式增长：我们生活在一个数据充斥的时代。从社交媒体的互动到智能设备的传感器，每天都有海量的数据产生。这些数据为数据科学家提供了丰富的“原料”，让我们能够从中提炼出有价值的洞见。

企业的实际需求：在各个行业，企业都意识到利用数据来获取竞争优势的必要性。无论是预测消费者行为、优化供应链，还是提升产品设计，数据科学的应用几乎是无所不在的。这种对数据驱动解决方案的需求，推动着数据科学的快速发展。

计算能力的飞速进步：计算技术的进步使得数据分析变得更加高效和强大。云计算、机器学习算法以及人工智能技术的发展，让数据科学家能够处理前所未有的复杂问题。

这三股力量不仅是数据科学发展的引擎，也是未来发展的关键所在。对于那些希望投身数据科学领域的学生和专业人士来说，我的建议是，密切关注数据、实际需求和计算技术的演进，这将是你们取得成功的关键。

数据科学与人工智能：紧密相连的未来

随着数据科学的发展，它与人工智能（AI）的关系也变得越来越紧密。人工智能依赖于海量数据来学习和决策，而数据科学正是这些数据的基础。可以说，数据科学的进步直接推动了人工智能的发展，造就了越来越智能和强大的系统。

对于希望进入这一领域的人来说，理解数据科学与人工智能的联系将至关重要。掌握处理大规模数据集、开发机器学习模型和应用人工智能技术的技能，将使你们在未来的职业生涯中占据有利位置。

给年轻一代的建议：为数据驱动的世界做好准备

数据科学沉静几十年而后急速发展的历史告诉我们，未来的发展不仅仅依赖于技术，还需要洞察力和适应力。对于考虑投身数据科学的年轻人，我想说以下几点建议：

关注数据趋势：时刻留意数据在不同行业中的产生和使用方式，这将帮助你们预测未来最有潜力的数据科学领域。

解决实际问题：数据科学的核心在于解决现实世界中的问题。培养解决问题的能力，将使你们在任何行业中都能发挥重要作用。

紧跟计算技术的进步：计算技术不断进步，保持对新算法和新硬件的了解，能够让你们在处理日益复杂的挑战时游刃有余。

培养跨学科能力：数据科学在多个学科的交汇处蓬勃发展。掌握统计学、编程以及特定领域的知识，将使你们成为更具竞争力的数据科学家。

结语：一个充满无限可能的领域

数据科学从一个不为人知的小众学科，逐渐发展成为今日的主流学科，足以证明它在当今世界中的重要性。随着数据的持续增长，企业需求的不断提升，以及计算能力的飞跃，数据科学的前景将更加光明。对于那些希望在这个领域中大展身手的人来说，了解并把握这些推动力将是未来成功的关键。通过不断学习和适应，你们将能够在这个充满挑战和机遇的领域中找到属于自己的位置。