

业界 | 尴尬了，数据故事讲不好，模型再酷炫都没用

大数据文摘 Yesterday



大数据文摘出品

编译：王一丁、Zhifu、什锦甜、夏雅薇

导言：数据科学风靡了几年，已经完成了从普及到应用的商业落地，越来越多的公司都已经同意数据驱动战略的重要性，但雇几个数据科学家和有一个数据团队，并不等同于公司就能坐享数据科学的果实。

数据科学家慢慢发现在科研阶段引以为豪的高端模型，在向没有技术背景的管理层做展示的时候毫无吸引力，得出的商业建议在付诸实施的时候也大打折扣。

作者认为要想做一个有影响力的数据科学家首先得学会讲好故事，他给出了三个有效建议来帮助数据科学家提高从洞察到行动的转换率。

有两类数据科学家：一类科学家专注于数据分析（A类），另一类专注于模型构建（B类）。

在数据科学项目中，两种科学家可能出现如下两种问题：

- 模型不适用于应用体系。
- 洞察的结果对于改进产品没有用处。

这两种结果常发生在从数据科学到其他领域的过渡之中：一种是从数据科学到软件工程，另一种是从数据科学到业务执行。

这两种转变是对现实生活创造价值和产生影响的关键，因此这个过程也存在很高的风险。

在模型交付过程中，软件工程师可能会说这些模型不符合他们的工程标准，商业部门的同事可能会说你的想法很有趣，但他们不会有所行动。这些这样那样的“借口”经常会丢给数据科学家们让他们很难取得成绩。这比数据科学领域其他挑战更令人担忧，因为我们无法左右数据分析成果在被转交后的命运。

尽管我们能做的非常有限，但我们应该采取一些行动来避免这两种结果。

模型部署在不同的机构中会有不同的效果，但是编写应用级代码可以帮助数据科学家进行更好的部署。

Trey Causey在他的博客中为数据科学家列出了编写应用级代码”所要做的一些事”

网址：

http://treycousey.com/software_dev_skills.html

简而言之就是，使代码模块化，创建文档，使用版本控制（包含但不限于详细记录和建立测试）。

许多数据科学家习惯使用自上而下的命令式编码；这种风格对与软件工程师合作制造了障碍。要使部署工作顺利进行，编写应用级代码很重要。

将想法转换为实际行动就像将“思维代码”部署到“社会机器”中一样。

如果数据科学家可以通过编写应用级代码来改进模型部署，那么数据科学家可以做些什么来改进商业战略落实呢？

用数据讲故事的能力是解决问题的关键。

数据背后引人入胜的故事可以给观众留下深刻印象并激励大家的行动力。用数据讲故事听起来很棒，但“讲故事”本身却是一个模糊的概念。为了真正深入了解数据故事，最近我回顾了

Cole Knaflie 撰写的《Storytelling with Data》和 Nathan Yau 编写的《Visualization that Means Something》。这两本书让我重新认识了用数据说话（讲故事）。

基于这两本书和我的经验，我想向你介绍优化结论向行动转换的策略。

概述：

- 一开始就将想法观点亮出来
- 在视觉层次上强调你的突破性结论
- 提升结论到行动的转化率

把观点性结论放在展示最开始的部分

几个月前，我学会了一种呈现数据科学结果的方法。我认为这种方法很有效，我看到很多科学论文都采用了这种方法。

所以我的展示流程是，我从背景开始，讨论数据源，使用图形和表格来突出数据属性，讨论模型构建，讨论建模结果，最后总结观点（适当时要求采取行动）。

在这个流程中，想法见解是放在最后的，因为对数据和模型的基本解释会让观众真正理解观点做适当铺垫。

这个流程对我很有用，然而对于我的商业合作伙伴却不是这样——这些合作伙伴是商业实操的资深人士，但却是数据科学领域的门外汉（如今大多数决策者还是没有受过数据科学的专业培训）。

他们对我的数据探索过程不好奇，对建模细节也不感兴趣。他们只对这些见解感兴趣，因此建议我将数据细节移到附录中。

尽管我惊讶于他们对科学推理过程的不重视，但在我采纳他们的建议后，发现我的数据讲故事能够更快地与他们产生共鸣。

看来隐去数理推理过程的确更受欢迎。

为了做得更好，我回过头重新审视那些与我观点不直接相关的内容。现在，我讲的故事更多地围绕着结论而不是数据建模建立的过程。

乔什·伯诺夫（Josh Bernoff）在《Writing Without Bullshit》这本书中的一段很好地概括了上述的这种现象：

你必须颠倒你在学校学到的关于推理和写作的知识。在学校里，你学会先热身，再从第一原理开始慢慢推导出结论。但商业背景的读者没有热身时间，对推理过程缺乏耐心，除非他们提前知道目的是什么，回报是什么。所以从鲜明的观点陈述和结论开始，然后配上你的依据和理由。这样，读者不必阅读整篇文档就可以有所收获。

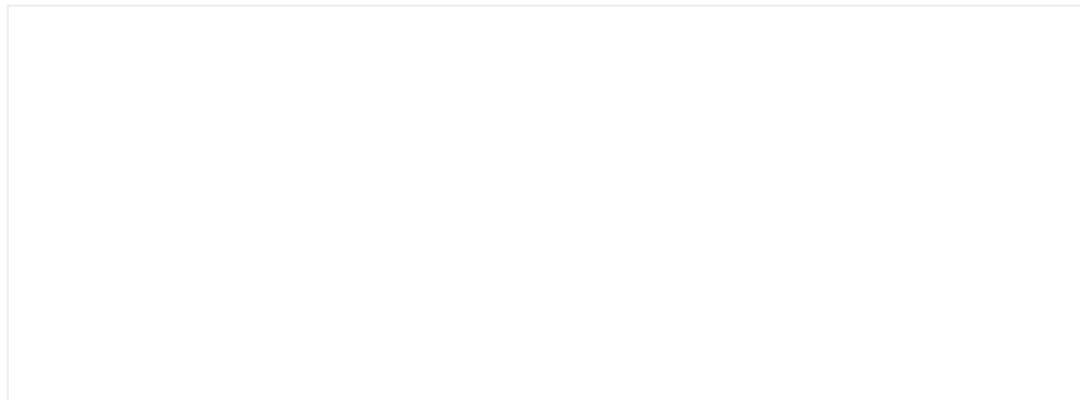
我仍然和我的技术同事用行话交流，然而，打交道的决策层次越高，我就越需要像Josh所说的那样先说结论再说依据。

另一个职业——咨询顾问，一直以来支撑着企业高管作出决策。

好像咨询顾问早就意识到需要先把观点摆在前面的重要性。

我找到了咨询顾问写的一篇关于家庭语音银行的研究论文。

本文采用的流程就是久经考验的讲故事方式。



来自咨询公司的研究论文中的目录

正如在上面的目录中可以看到的，顾问在“执行摘要”之后需立即提出“建议”，而将研究方法放在最后一个章节。

其中在一开始就有大胆的结论和建议，也包含了决策可能产生的后果。

在本文中，市场对智能语音银行业务的需求很低，因此顾问建议银行对此持谨慎态度。对于顾问而言这是常见的方法。无论是写作还是演讲，在最显眼的位置需要展示最重要的内容。

为了实现见解与行动的转化，应该把观点放在最开始的地方。

在视觉层次上强调突出你的结论倾向

数据可视化效果直接决定数据故事的好坏。数据可视化有表格和图形，包括视觉线索，坐标系，比例尺，周围文本等形式。

数据可视化既是一种很好的信息交流方式，所谓“一图胜千言”；同时也是一种很好的分析方式，正如John Tukey在《探索性数据分析（Exploratory Data Analysis）》中所说的“图片的最大价值在于它使我们注意到我们之前从未注意过的东西”。

鉴于数据可视化的有用性，我们可以通过更好地使用数据可视化来显著提高数据的“故事性”。

以前我的数据可视化优化策略是把图表做得更好看。我会浏览我起草的表格和图表，选择最有趣的表格，添加缺失的标题、图例和颜色，最后将它们放入我的演示文稿中。

这种方法一直效果不错，直到最近，当信息就呈现在我的同事面前，但他们还在问我信息在哪儿的时候。我这才意识到，我的图表虽然展示了信息但并没有传达出该关注图表哪个区域的信号。

在《Storytelling with Data and Visualization that Means Something》一书中，借助潜意识属性来创建有层次的视觉结构这一概念可以帮助我们体验到数据可视化真正的力量。

数据可视化不仅仅是使用正确的格式展示数据，更是用暗示性的视觉指示来引导观众更好地理解图片给出的信息，在视觉结构的最顶层给出观点。

我将使用两个案例进行说明。

以下是Cole Knafllic的案例研究。下面的第一张图是典型的线性图。线形图多用于在时间序列上的研究、重复测量研究等。

下面的图表有正确的标题、坐标轴和图例。从技术标准来看这是一个还不错的图表。但是，读者几乎不知道它想说什么。



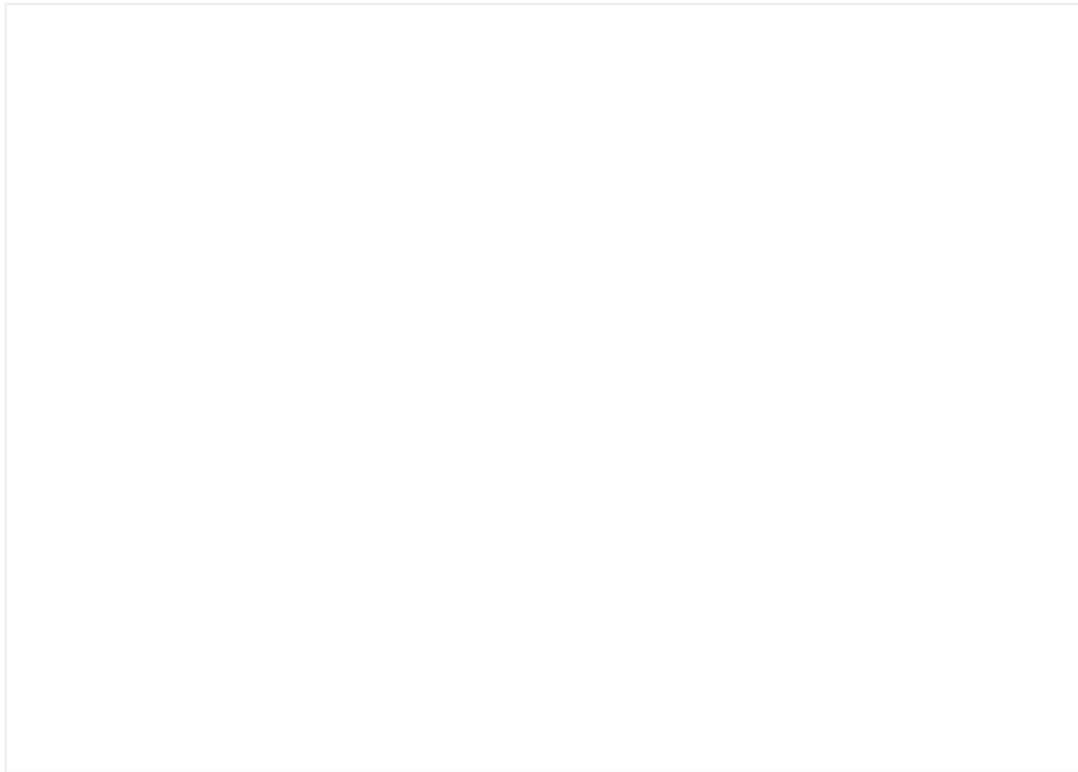
数据是由投资者提供的，百分比之和大于100是因为调查者可以作出多个选择

改进之后的版本则是使用了潜意识属性。下面的图表中使用加粗的蓝色折线突出了最大的增长趋势，而其它所有的趋势则是采用了灰色，使其和背景融为一体。经由更好的视觉层次的结构，读者可以更加轻松的理解这张数据可视化图表。



在我采取这种方法之后，每当我采用前注意的特性来突出我想表达的内容的时候，我可以立刻注意到观众的反馈来得更快，更加强烈而且更加积极。这证明了在视觉层次的强调的确有效果。

在另外一个案例研究中，Cole Knafllic阐述了标题的强大之处。在下面的图表中，标题不是典型的“请求数量的时间趋势图”，这样的标题是描述性的，但不够直接。而这里给出的标题是“请批准雇佣两名全职员工”，给出了一个明确可采取的措施。这个建议帮助读者从趋势线中找到了因为两名员工的离职所导致的绩效差距。



我总是使用描述性的标题：“X随Y变化的趋势”，“A随B的分布”等等。现在我的标题是一个建议，问题或者真知灼见。标题是数据可视化中最重要的一个部分，一个充满张力的标题将极大提升信息输出效率。

我强烈推荐Cole和Nathan的图书《Storytelling with Data: A Data Visualization Guide for Business Professionals》，书中还有很多其它的技巧会对你产生帮助。

链接：

<http://flowingdata.com>

<http://www.storytellingwithdata.com/>

提高洞察到行动的转化率

利用数据讲故事不是一个一次性的展示。你会被要求一遍又一遍的对不同利益相关者讲述同一个故事，特别是当你的真知灼见成功地与初始受众产生共鸣的时候。整个过程就像一个路演，你可以利用你的数据故事来活得尽可能多的支持。

有了足够多的支持之后，人们就会开始采取行动：利用不同的方式定位消费潜力，调整目前的战略从而提高运营效率，关闭不必要的产品功能，推出新流程来防范风险等等。

正如中国的一句谚语：牵一发而动全身。把你得出的建议转化为行动时，会对组织的各个部分产生连锁反应，而每个部门自然会从多个方向进行审查，而这些审查对于你的分析来说相当于一次压力测试。

本着一种怀疑的态度，人们可以对你的数据质量，假设和建模方法提出质疑。如果人们在你的原始数据中发现了错误，那么你故事的说服力会降低。

从不同的角度来看，人们可能会在你的关注点之外提出各种问题，这可能会导致额外的分析迭代过程。即使你基于数据的故事被别人完全接受，人们也可能会要求你提供数据，图表和数据模型从而进行进一步的分析。

人们还可能会要求你为了其它会议提供不同的图表。这些需求将促使你将分析的工作文档记录在案，使其可访问，模块化和可交互。如果你通过了这些压力测试并不断改进你的故事，你的故事会变得更加精巧，并且与实际行动更加接近。

路演有可能会成功，失败，或者介乎二者之间。A型数据科学家需要对预期和实际行动之间的转化情况有合理的预估。时至今日，每个商业领袖都高度重视由数据驱动的决策，而给出数据解决方案成为商业世界中最重要的一环。

但是这种趋势并不意味着决策仅仅将由数据驱动。决策本身仍然是科学与艺术的复杂调和，是数据逻辑和业务逻辑的混合体。在现实世界中，高管利用群策群力，定性和定量地来做出决策。

数据科学分析能够完全决定一项主要商业活动这句话并没有错误，但是也可能会起到误导作用。在过去，我认为如果一个数据分析的结果最终导致一个商业行动，那么这个数据分析本

身是成功的。但是我现在看到了更多的商业案例，在这些案例中我们获得了一些真知灼见，但并没有转化为行动。

通过研究复杂的商业世界，我现在得出的结论是：将商业洞察转化为行动还取决于你的洞察有多大的影响力。而且A型数据科学家的成功不应该由它们的分析所带来的商业行动（和业务成果）来判断。放下对于数据科学可以化腐朽为神奇的幻想，并且承认洞察力的实际局限性有助于让我更好地专注打磨技巧，而不是被许多无法控制的事情分散注意力。

最后想说的话

商业洞察是数据科学工作的重要输出。我们寻找数据洞察力，因为我们需要洞察力来帮助我们让世界变得更美好。将洞察力转化为行动是将有价值的洞察力变为现实的方式。

由于数据科学家通常会将分析结果交给其他专业人士作决策，数据科学家应该更好地讲述数据背后的故事，从而能够更好地将洞察力转化为行动。通过把将商业建议放到演讲最开始的部分，并且在视觉上着重强调，数据科学家可以讲述更加打动人的数据故事，从而保持一个合理的从洞察到行动的转化率。

相关报道：

<https://towardsdatascience.com/data-storytelling-improve-insight-to-action-conversion-for-a-greater-real-world-impact-b94e1827fa01>

【今日机器学习概念】