

# 数据故事化：从数据感知到数据认知

朝乐门 张 晨

**摘 要** 数据呈现是大数据时代的新课题。通常,数据呈现的主要途经有两个,即数据的可视化和数据的故事化,二者的区别在于,数据可视化主要解决的是数据感知问题,而数据故事化更加关注的是如何将数据感知转换为数据认知。数据故事化涉及三个基本要素:数据、视觉效果和叙述。从数据故事的创作者和受众之间的信息交流模式看,可将数据故事化分为创作者驱动和受众驱动两种不同模式。目前,数据故事化中常用的结构有三种:马提尼酒杯结构、互动演示幻灯结构和向下钻取事结构。数据故事化的主要活动包括理解数据、明确目的、了解受众、确定关键数据、选择故事模型以及故事叙述。数据故事化是数据科学的主要研究内容之一,也是数据科学区别于其他学科的重要特征。图6。表4。参考文献60。

**关键词** 数据科学 数据故事化 数据可视化 数据感知 数据认知

**分类号** TP393

## Data Storytelling: From Data Perception to Data Cognition

CHAO Lemen & ZHANG Chen

### ABSTRACT

There are two main methods for data presentation: data visualization and data storytelling. The difference between them is that data visualization mainly solves the problem of data perception, while data storytelling focuses more on how to transform data perception into data cognition.

Data storytelling involves three basic elements: data, visuals, and narrative. Combining data and visuals can enlighten the audience to gain data insights; applying narrative into data can explain the data phenomena; coupling visuals with narrative can entertain the audience. The main activity flow of data storytelling is divided into six steps: 1) The author must understand the data and make clear the meaning it represents. 2) The author who makes data stories needs to have a clear purpose. 3) Identifying the audiences as novices, generalists, managers, experts or executives helps author create different data stories for different audiences. 4) The next step is that author identifies key data and uses the most effective data to describe the data story. 5) Then the author chooses the story model, chooses the appropriate chart, and describes the information presented by the story according to people's visual characteristics. 6) Providing background narration and guiding the audience according to the plot effectively synthesizes and organizes data stories. In addition, the following three special activities may be involved in the actual data storytelling project: 1) Data story-telling experiment and pre-investigation. 2) Continuous improvement of data storytelling. 3) Separation between the author and the narrator of the data story.

通信作者:张晨,Email:zhchen009@ruc.edu.cn,ORCID:0000-0001-8218-0348( Correspondence should be addressed to Zhang Chen, zhchen009@ruc.edu.cn,ORCID:0000-0001-8218-0348)

It is worth mentioning that the transformation from data to story and the presentation of data story are two different stages of data storytelling. The author of data story needs to convert the data into a story model beforehand, and then the narrator takes a special way to present the story. There are two presentation forms of data stories: 1) The narrator tells the story and the audience listens to the story. 2) The narrator shows the data story to the audience, and the audience can see the story.

Data storytelling is one of the main research contents of data science, and it is also an important feature that distinguishes data science from other disciplines. The application of data storytelling in data science is mainly reflected in four aspects. First, data storytelling solves the “last mile” problem of data science, which plays a crucial role in the success of data science projects; second, data storytelling is an important means of obtaining insights from big data; third, data storytelling can transform data insights into data actions; fourth, data storytelling is an important activity of data product development, such as data journalism.

Finally, this paper summarizes five main characteristics of the current research: 1) Foreign research is more than domestic research. 2) Research articles published informally (such as blogs) are more than the officially published academic papers. 3) Theoretical research lags behind the practice application. 4) There is more research on concept level than the realization of concrete technology. 5) The visualization tools for data story are more than specialized tools of data storytelling. In view of these characteristics, some recommended topics in the following research of data storytelling are proposed: 1) Improving the theoretical system of data storytelling. 2) Studying the evaluation method of data storytelling, optimizing specific data story projects. 3) Strengthening interdisciplinary research and further broadening the research perspective and theoretical basis of data storytelling. 4) Exploring new algorithms and models for data storytelling. 5) Developing special tools for data storytelling. 6 figs. 4 tabs. 60 refs.

## KEY WORDS

Data science. Data storytelling. Data visualization. Data perception. Data cognition.

## 0 引言

作为大数据时代的新兴活动,数据的故事化呈现已经成为数据科学及相关研究领域的热门话题之一。有研究调查发现,能够记住“故事”的人数可以达到63%,但是能够记住孤立的统计数据的人数只有5%<sup>[1]</sup>。在数据科学中,数据的可视化和故事化具有不可替代的作用。数据可视化具有易于理解、易于感知和易于洞察的特点,而数据故事化具有易于记忆、易于认知和易于体验的特点<sup>[2]</sup>。例如,在一项“拯救孩子们(Save the Children)”的公益活动中,研究者为同一个被捐款者准备两种不同版本的宣传手

册,一种是基于故事化描述的版本,另一种是简单罗列事实数据的版本,结果发现拿到前一版本的捐赠者的平均捐款金额比后者高出两倍以上<sup>[3]</sup>。

从主体的数据接受模式看,感知是认知的前提,认知是感知的延续,数据可视化和数据故事化分别解决的是数据的感知和认知问题。因此,在大数据时代,数据的可视化和故事化通常交叉应用,先采用可视化方式引起人们的感知活动,然后通过故事化方式达到进一步认知的目的。所以,也有人认为数据故事化是数据可视化处理的下一步<sup>[4]</sup>。

数据故事化作为数据科学的“最后一英里”问题<sup>[5]</sup>,其重要性往往容易受到数据科学家的

忽视。因此,本文从数据故事的模型与结构、数据故事化的流程与活动、数据故事的类型与呈现、数据故事化对数据科学的意义等四个方面对数据故事化的理论研究现状及具体应用情况进行梳理,这对于实现数据价值的最大化有重要的理论意义与实践意义。

## 1 相关研究

### 1.1 数据故事化的定义

数据故事化(Storytelling)是指为了提升数据的可理解性、可记忆性及可体验性,将“数据”还原或关联至特定情景,并以叙述方式呈现的过程。数据故事化也是数据转换的表现形式之一,其本质是以“故事叙述”的方式呈现“从数据中发现的洞察”。数据故事化中的情景,可以是还原情景、移植情景和虚构情景<sup>[2]</sup>。

数据故事化的概念和提法有很多,如数据驱动型故事化、可视故事化、分析型故事化、交互式故事化和用数据讲故事等,从不同视角或层次讨论了数据故事化的某个(些)内容。其中,数据驱动型故事化(Data-Driven Storytelling)<sup>[6]</sup>强调的是故事叙述的一种形式,区别于模型驱动和目标驱动的叙述方式;可视故事化(Visual Storytelling)<sup>[7]</sup>主要强调的是故事叙述中可视化呈现的重要性;分析型故事化(Alytical Storytelling)<sup>[8]</sup>则强调的是数据分析在数据故事化中的重要地位,数据故事化并非是简单地将数据改编成故事,而是在对数据进行分析挖掘的基础上,对数据背后的潜在模式信息进行故事化;交互式故事化(Interactive Storytelling)<sup>[9]</sup>主要强调的是受众交互在数据可视化中的重要地位,数据可视化是一个交互式创造过程<sup>[10]</sup>;用数据讲故事(Storytelling with data)<sup>[11]</sup>主要强调的是故事的客观性,突出故事要有依据或凭证,避免故事变得过于主观。

数据故事化旨在从故事的维度将繁琐复杂的数据进行图标化的整合,从提炼分析出的数据中组织出一个故事,从而帮助决策者做出基

于数据分析的决定<sup>[5]</sup>。数据故事化的结果是针对特定的数据生成一个通俗易懂的数据故事,数据故事通过一系列链接的可视化来探索和解释数据随时间或某种因素变化的方式和原因<sup>[12]</sup>。数据故事的叙述者要根据数据类型以及传递给受众的数据来选择合适的可视化方式,并将这些可视化结果整合进时间或概念序列中,从而塑造成一个叙述故事,以帮助揭示一些洞察、趋势或潜在模式。

数据故事化的目的是使数据更易于认知,帮助受众从数据中提取洞察,鼓励和激励批判性思维,以便做出过程分析、业务决策或提出讨论。数据故事化的基本原则包括忠于原始数据原则、设定共同情景原则、体验式叙述原则、个性化定制原则、有效性利用原则和3C精神原则<sup>[2]</sup>。

### 1.2 数据故事化的研究内容

数据故事化已成为大数据时代的热门话题之一,其理论研究主要集中在以下几个方面。

#### (1) 数据故事化的理论基础

在现有研究中,数据故事化的理论基础涉及认知科学、信息图形学、可视分析学和数据科学。①认知科学:基于认知科学的数据故事化研究对人类学习的本质提出三个主要假设,即双通道假设、有限容量假设和主动学习假设<sup>[13]</sup>;②信息图形学(Information Graphic):数据故事化依赖于对视觉信息进行有效设计,通过插入文字注释和图表呈现等形式传递抽象、复杂和密集的信息<sup>[14]</sup>;③可视分析学:数据故事是数据可视化进一步发展的形式,当故事中包含大量数据时,数据可视化可作为讲故事的交流媒介<sup>[4]</sup>。④数据科学:数据科学的理论,尤其是数据产品开发理论为数据故事化提供了理论基础<sup>[2]</sup>。

#### (2) 数据故事化的方法研究

在现有研究中,数据故事化主要采用作者目标驱动、受众行为驱动、视觉增强和数据注释等方法。①作者目标驱动方法,即建立明确的目标,并根据目标再进行数据故事描述<sup>[15]</sup>;②受众行为驱动方法,即采取扫描、翻转、交换和向

下钻取等方式实现数据故事化<sup>[16]</sup>;③视觉增强方法,通过可视化方法来增强数据故事的视觉效果和受众体验;④数据注释方法,利用文本注释帮助受众理解信息图形,以及数据与业务之间的关联<sup>[17]</sup>。

### (3) 数据故事化的应用场景

就现有研究而言,数据故事化的主要应用场景涉及数据新闻、商务演示和学术交流等领域。①数据新闻:数据故事(Data Stories)或用数据讲故事(Storytelling with Data)已成为新闻业的新趋势之一<sup>[18]</sup>。例如,2017年,Google和PolicyViz合作,对美国、英国、德国和法国的记者进行了56次面谈,并对900多名记者进行在线调查,结果表明42%的记者定期使用数据讲故事<sup>[19]</sup>。②商务演示:数据故事化在商业领域主要

用于与故事受众进行交流和互动,主要体现在协助公司内部人员做出业务决策<sup>[20]</sup>、向客户解释复杂的数据点和趋势<sup>[21]</sup>、提升品牌知名度、进行有效营销<sup>[22]</sup>等。③学术交流:在学术界,用数据讲故事不仅适用于内部讨论和决策制定,而且还应用在流程分析<sup>[4]</sup>。数据故事化在学术领域的应用更多地体现在学者之间的面对面沟通和会议交流。

### (4) 数据故事化的软件工具

目前,支持数据故事化功能的软件工具越来越多,比较有代表性的是TIBCO Spotfire、Tableau、Power BI、Flexdashboard、D3.js、Qlik Sense、Banjo和Narratives for Tableau等(见表1),可分为两大类,即支持数据故事化的可视化工具和专门用于数据故事化的工具。

表1 支持数据故事化功能的软件工具

工具名称	主要特征	类别
TIBCO Spotfire	①应用于商业领域,使用主体往往是企业中从事销售、技术和科研工作的专家; ②为决策提供便利,将决策者从复杂的统计模型中解放出来,不再需要关注模型中复杂的或细节性知识,而是从应用角度理解和运用模型,有效降低了决策的不确定性。	支持数据故事化的可视化工具
Tableau	①支持7种类型的数据故事; ②易于使用,采用拖拽的方式进行操作。	
Power BI	①主要用于商业领域,支持云计算业务分析; ②输出一个可共享的在线报告,可自动更新以提供实时洞察。	
Flexdashboard	①基于R; ②布局灵活,故事板布局可用于呈现可视化序列和相关评论。	
D3.js	①基于JavaScript; ②依赖于标准的Web技术来绘制可视化元素,如HTML、SVG、CSS。	
Qlik Sense	①主要用于商业领域,可扩展性好,提供用户自定义控件编程; ②易于使用,采用拖拽的方式进行操作。	
Banjo	①主要应用于新闻领域,允许用户实时体验突发新闻和现场活动; ②数据来源广泛,从超过12亿个公共社交媒体账户中收集信息。	专门用于数据故事化的工具
Narratives for Tableau	①基于Tableau Server、Tableau Online、Tableau Public; ②自动为交互部分的数据添加故事叙述。	

### 1.3 数据故事化的研究热点与趋势

通过国内外文献调研发现,数据故事化的

当前研究热点和未来发展势如下。

(1) 数据故事化的研究热点问题包括以下

几个方面。①关于数据故事化的重要性的讨论,数据故事化可以使数据证据和分析结果更具说服力<sup>[23]</sup>;数据故事化是每个数据人才都必须掌握的重要技能<sup>[24]</sup>。②数据故事的基本模型与结构,如 Segel 和 Heer 提出三种常见的数据故事化模型<sup>[25]</sup>;Aristotle 提出五步叙述结构<sup>[26]</sup>;成熟度曲线模型<sup>[27]</sup>;Zawadzki 提出数据故事化的 SPSN 框架和 SUCCEs 模型<sup>[28]</sup>。③数据故事化的流程:Chibana 将数据故事化分为五个步骤,并重点关注了数据故事化过程中图表的设计流程<sup>[29]</sup>;Taborga 强调角色驱动在数据故事化中的重要性,提出数据故事化包括定义行动号召、选择受众、保证数据透明、使用简单的术语、使用可视化来补充叙述和制作真实的故事六个步骤<sup>[30]</sup>;Maptive 公司从新闻叙事的视角提出制作引人注目的数据故事的六个步骤:提出一个好问题、区分受众类型、查找数据、检查数据源、过滤数据分析结果、确定数据可视化方式<sup>[31]</sup>。④数据故事的类型:Davenport 从时间维度、关注点维度、深度维度和方法维度将数据故事划分为不同的类型<sup>[32]</sup>;Tableau 从场景视角划分出七种不同类型的数据故事<sup>[33]</sup>;Proschool 从目的视角把数据故事划分为说明性数据故事、解释性数据故事和预测性数据故事<sup>[34]</sup>。⑤数据可视化的设计:Knaflie 提出要遵循理解上下文、选择合适图表、消除杂乱、引导受众注意、像设计师一样思考等原则<sup>[11]</sup>;HubSpot 建议消除过多标签,以使用可视化的方式比较数据值,不要在同一界面使用超过六种颜色,不要使用 3D 图表<sup>[35]</sup>;Evergreen、Ann Emery 从布局、配色和排版三个方面给出设计方案<sup>[36]</sup>。

(2)数据故事化研究的主要发展趋势体现在三个方面。①数据故事的智能化生成:数据故事化的发展将经历全人工的数据故事化、半自动化的数据故事化以及全自动化的数据故事化等演变过程<sup>[37]</sup>;自动生成数据故事<sup>[38]</sup>,实时生成数据故事,人人都可以是数据故事的叙述者,是数据故事化的主要发展趋势<sup>[39]</sup>。②数据故事化在数据科学的广泛应用,主要体现在使

数据可视化更有效<sup>[40]</sup>、充分发挥数据价值<sup>[41]</sup>、提高数据科学家的业务能力<sup>[42]</sup>以及开发高质量数据产品<sup>[43]</sup>。③加强数据故事化本身的理论研究和实践:目前,数据故事化虽然得到了广泛关注,但是针对其本身的系统研究并不深入。数据故事化理论的进一步发展及相关实践的可持续发展应用亟待加强对其本身的系统研究。

## 2 数据故事的模型与结构

Dykes 认为,数据的故事呈现涉及三个重要因素:数据、视觉效果和叙述<sup>[24]</sup>,如图 1 所示。数据故事经常采用数据呈现中的数据可视化技术,但它涉及的不仅仅是呈现图形,还要逐步引导受众认识数据、了解数据并得出结论。将数据和视觉效果结合起来,可以形成数据的可视化展示,帮助他人加强对数据的理解,得到从数据集中难以发现的认识;将叙述应用于数据,可以解释数据中正在发生的事情,以及背后隐含的信息;将视觉效果和叙述相结合有助于吸引他人的注意,以受众更容易接收的方式让他们参与其中,产生共鸣。而有价值的数 据、效果最佳的视觉展示和合理的叙述融合到一起时,将产生有影响力的数据故事。此外,数据故事化过程需要创作者和受众的参与,创作者描述出令人印象深刻的数据故事,并把它展示给受众;受众接收数据故事,并采取相应的行动。

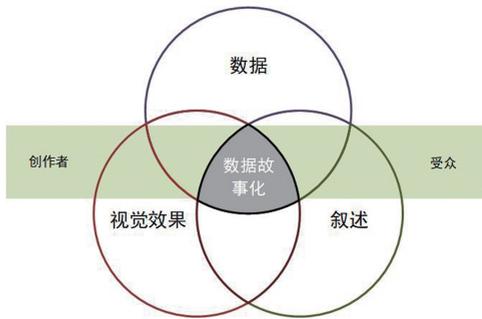


图 1 数据故事化要素

## 2.1 数据故事的模型

数据故事化依赖于数据故事的创作者和受众之间的互动和信息传递,而好的数据故事必须在创作者驱动和受众驱动之间保持平衡,创作者既要为受众提供结构化的故事叙述又要与其保持一定的交互空间。Segel 和 Heer 提出数据故事有创作者驱动和受众驱动两种形式(见表 2),并给出三种常见的数据故事化模型:马提尼酒杯(Martini Glass)结构、互动演示幻灯结构(Interactive Slideshow)和向下钻取事(Drill-Down Story)结构<sup>[25]</sup>。

表 2 创作者驱动和受众驱动的数据故事化

内容	创作者驱动	受众驱动
流程	线性流程	非线性流程
信息量	固定	不固定
与受众交互	少	多

创作者驱动的数据故事是指在创作者进行故事化描述和呈现的过程中以线性路径为主,不允许受众与图表交互,数据和可视化是由创作者选择的,并作为成品呈现给读者,重在向受众进行数据传递,是单向地将故事化的数据展示给受众并让受众接纳的方法。当讲故事是为了进行有效沟通或让他人接纳确定性的信息时,通常采用这种方法,例如在电影、杂志文章、广告、商业演示中就以创作者驱动为主。

受众驱动的数据故事是强调受众的高度参与,并为受众提供处理数据的方法。创作者不是对受众进行简单的信息传达,也不依赖严格的结构化故事框架,创作者只负责提供数据及其可视化方式,而由受众参与其中,对可视化的图形进行架构,形成受众自己的故事流。例如,在 Tableau 等可视化分析工具中,读者驱动的方法支持数据诊断、模式发现和假设形成等任务。

(1) 马提尼酒杯结构是最常见的数据故事化形式,长长的杯颈代表单一路径的创作者叙述,大开口的杯口代表读者参与的交互,这是一种先是创作者驱动进行叙述,再是受众驱动进

行互动的数据故事化模型。在这个过程中,创作者先根据自己对数据的分析,整理出一个完整的数据故事,然后由叙述者向受众叙述这个预期的故事,传达创作者的观点并为受众提供观察视角和讨论焦点,再让受众针对已有的数据故事发起讨论,自由探索问题。

(2) 互动演示幻灯结构是在叙述过程中创作者和受众共同参与的一种方式。在叙述完故事的每一个发展阶段后,受众可以提出问题并进行讨论,但对每一个发展阶段的叙述过程中仍然遵循马提尼酒杯结构。这种方式可以帮助受众更好地跟进创作者向他们传递的意图,也可以在必要的时候重复叙述数据故事的步骤,以对复杂数据集和数据故事的形成过程有更深入的理解。

(3) 向下钻取事结构是以受众驱动为主,为受众呈现出数据可视化的结果页面,受众可以在页面上操作选择,以查看特定的细节和故事发展趋势。在这种故事模式的主导下,创作者允许受众深入其中选择想要查看的故事流,由受众自己决定什么时候了解什么样的故事。

## 2.2 数据故事的结构

与通常意义上的故事一样,数据故事也会有开头、中间和结尾。数据故事的结尾不是一个固定的事件,而是通过一系列选项或问题来触发受众的行动<sup>[44]</sup>。在数据故事的结构方面,比较有代表性的是 Aristotle 提出的五步叙述结构<sup>[26]</sup>和成熟度曲线模型<sup>[27]</sup>。

(1) 五步叙述结构,即数据故事的主要结构包括以下 5 个基本步骤:①提供一个能引起受众兴趣的故事或陈述;②提出需要解决的问题或需要回答的疑问;③为提出的问题提供解决方案;④描述在解决方案中采取行动的具体好处;⑤发出行动号召。

(2) 成熟度曲线模型,即数据故事的结构包括引言、上升期、高潮、下降期和结局五个部分(见图 2):①引言:在第一部分要向受众介绍数据故事的背景(如数据源、时间轴、以前的趋势等);②上升期:介绍当前遇到的问题,以及数据

的变化趋势;③高潮:让受众完全了解问题现状并理解数据的含义;④下降期:进行决策分析或提出讨论,激励受众采取行动;⑤结局:利用数据解决问题。比较有代表性的是 Gartner 通常采

用成长曲线(Hype Cycle)来叙述新兴技术领域的发呈现状与趋势,如 Gartner 的数据科学本身的成长曲线(Hype Cycle for Data Science)<sup>[45]</sup>。

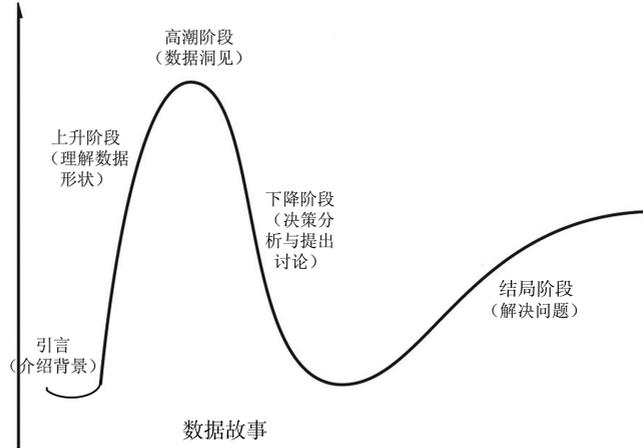


图 2 成熟度曲线模型<sup>[46]</sup>

### 3 数据故事化的流程与活动

数据故事化不涉及前期的数据收集、整理、分析等工作,而是将重点放在数据分析结果的叙

述性可视化呈现上。数据故事的描述是将数据以易于记忆、易于认知、易于体验的方式整合进故事流中,从流程来看,主要包括理解数据、明确目的、了解受众、识别关键数据、选择故事模型与呈现方式、故事叙述 6 个基本活动,如图 3 所示。

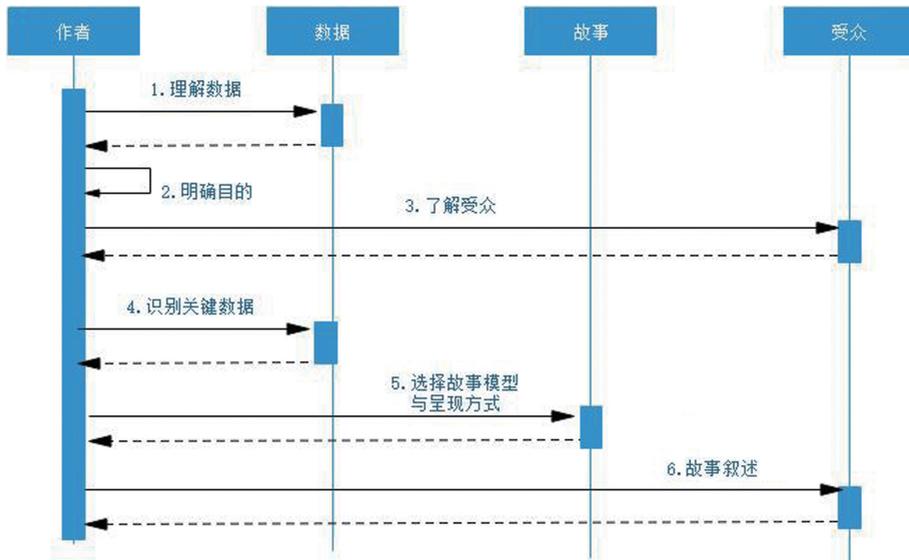


图 3 数据故事化的基本流程

### 3.1 理解数据

理解数据是数据故事化的第一步。理解数据也是数据故事化与通常意义上的故事的区别之一。在数据故事化中,可以通过探索型数据分析(Exploratory Data Analysis, EDA)等统计学或机器学习的方法以及与数据所属领域专家沟通等方法达到理解数据目的。在传统数据工作中,很多数据工作者习惯于采用统计学方法和机器学习方法来理解数据,而忽略数据可视化在理解数据中的重要意义。以统计学家 Anscombe 的四组数据(Anscombe's Quartet)为例,他于1973年提出了四组统计特征基本相同的数据集<sup>[47]</sup>,从统计学角度看难以找出其中的区别,但可视化后很容易找出。可见,探索性分析方法在数据故事化的第一步具有不可替代的重要地位。

### 3.2 明确目的

数据故事化要有明确的目的,而目的往往由数据故事的创作者或数据故事化项目的委托方决定。数据故事化的目的可以分为两个层面,一个是业务层面,另一个是科学层面。从业务层面看,数据故事化需要满足具体业务需求,而业务需求由数据故事化项目的委托方决定,具体可以分为描述性数据故事化、诊断性数据故事化、预测性数据故事化和规范性数据故事化(Prescriptive Data Storytelling)。从科学层面看,数据故事化的目的是弥补其他数据呈现方法,尤其是数据可视化方法的局限性,提升数据在受众中的可记忆性,需要突显数据故事的记忆难度和记忆留存时间。

当然,在确定数据故事化的目的时需要遵循 SMART 原则<sup>[48]</sup>,注意目的的具体(Specific)、可测量(Measurable)、可实现(Attainable)、其他目标具有相关性(Relevant)以及具有明确的截止期限(Time-bound)。此外,数据故事化应聚焦于有限目标,有限目标是数据故事化中必须重视的问题。

### 3.3 了解受众

数据故事化实际上也是数据价值的传递及

再增值过程,数据分析的结果最终都要传达给特定的受众,再由受众与自己的业务整合后采取进一步的行动。因此数据故事化的创作者需要根据受众的业务范围、知识背景、能力以及目的设定相应的故事背景,定制不同的数据故事,以方便受众理解数据。

戴尔公司执行策略师 Stikeleather 将受众分为五个主要群体,即新手、通才、管理层、专家和执行官<sup>[49]</sup>。新手第一次接触这个业务主题,但不希望业务过于简单化;通才对当前的业务主题有一定的了解,但需要搞清楚总体概述和主要问题;管理层试图对错综复杂的事物以及其中细节之间的相互关系有深入、可操作性的理解;专家希望能对研究主题有更多的探索 and 发现,而不是过多地谈论细节;执行官则只需要知道加权概率的意义和结论。对每一个受众进行细分,在了解受众的需求之后针对不同受众采取不同的叙述数据故事的方式和内容,避免过于泛化的受众,这样才能让受众准确理解数据的价值并与之产生共鸣。

### 3.4 识别关键数据

当前我们可接触到的数据是海量的,从这些海量数据中可以分析得到大量不同的结果,但过于冗余的数据往往会分散受众的注意力,无法从中提取出最有效的信息。为避免繁杂无序的信息给受众带来困扰,数据故事化要突出显示最重要的内容,这些关键数据需要满足以下原则:①数据客观、有效;②与受众的业务主题直接相关;③与创作者想要传达的意图直接相关;④不流于表面,数据背后蕴含着更深层意义的价值,可以解释某种现象的原因或揭示接下来的发展趋势。

### 3.5 选择故事模型

在理解数据、明确数据故事化的目的、了解目标受众、识别关键数据之后,需要对数据进行故事化建模。从前文的讨论可以看出,数据故事化分为创作者驱动和受众驱动两种不同模式

以及马提尼酒杯结构、互动演示幻灯结构和向下钻取事结构等常用结构。然而,故事模型的选择中需要注意的是视觉通道的选择。受众喜欢视觉元素而非演示文稿中的数字,如果以可视化形式显示,则会使他们更准确地记住信息。无论是利益相关者还是客户,合理的可视化都会对受众产生重要的影响。

(1)选择合适的图表。可以把数据分析结果呈现出来的图表形式有很多,例如词云、表格、点图、线图、条形图、饼图等。如果采用不恰当的图表则会为受众的认知增加负担。每种图表都有其最佳的使用场景,在利用图表可视化

数据时,应考虑到它们各自的优势与使用条件,一些常见图表的适用场景和适用数据如表3所示。条形图和线图简单图表比较常见,但有的创作者为了展示自己的可视化技术或者为了制造更加丰富绚丽的视觉效果而刻意避免使用这类简单的图表。实际上,正是因为这类图表是大多数人都熟悉的,受众在看到此类图形时不必费力去理解图形的结构以及图形形状代表的含义,继而可以把注意力集中在图表所传递的信息上。因此在数据故事化描述时简单图表的效果往往比华丽的复杂图表更好,这类容易阅读的简单图表应该成为创作者优先选择的对象。

表3 图表类型及适用条件

图表类型	适用场景	适用数据
词云	突出单词的使用频率,了解文本的中心内容	文本数据
表格	显示精确值、显示不同度量单位的数值、在整体数据中比较个别数据	数字数据
点图	显示不同变量数据之间的关系	混合数据
线图	反映数据变化的趋势	随时间变化的数据
条形图	比较不同数据之间的差别	混合数据
饼图	显示数据之间各自所占的比例情况	混合数据
雷达图	比较同类别的不同属性或不同类别的相同属性之间的情况	多维数据
地图	在地图上显示特定位置和区域	与特定位置和区域相关的数据

(2)按照视觉特点进行描述。人的视野非常有限,无法一眼获取到所有的信息,在视觉认知的过程中,人的大脑获取信息的方式与眼球的移动有关,人的大脑会随着眼睛的移动而获取各式各样的信息,因此必须将信息进行有效组织,按视觉顺序进行编排。人在无意识时视觉顺序遵循一定的模式:自上而下、从左至右的阅读顺序、关注突出信息以及视线容易受干扰。根据这种模式来更有效地可视化数据,可以极大地提高受众从中获取数据洞察的效率。人的视觉阅读习惯取决于经验,我们的各种出版物都是自上而下、从左至右编排的,因此养成的视觉习惯也与此顺序相同,形成由版面的左上方

开始,平行移动到右上方后,往下至左下方再后移至右下方的“Z字形”移动路线,如图4所示。图表的有效信息应按照此顺序排列,逐步引导受众视线。

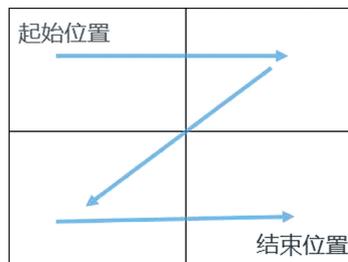


图4 多数人的视觉习惯

遵循这样的原则,图表的主题介绍、结论或者标题放在左上角会让受众在一开始就了解图表的主要内容,带着这样的心理预期解读接下来的图表会取得更好的效果。值得注意的是,反映趋势的图表竖直放置更加符合视线的水平移动,比水平放置效果更好。

受众在审视图表时,注意力会被一些格外突出的信息吸引,适当地突出想要向受众强调的数据信息,会提高数据故事的阅读效率,这些被突出的信息也称为前注意属性。前注意属性包括方向、形状、线条长度、线条宽度等<sup>[11]</sup>(见图5),我们的大脑不需要任何有意识的思考就能获取这一信息。

前注意属性在两方面非常有用:快速引导受众的注意力到你希望的地方;建立信息的视觉层次<sup>[11]</sup>。前注意属性即在版面中格外突出的信息表示,能够引导受众的视线,快速把受众的注意力吸引到创作者期望呈现的内容上;同时,突出显示的信息与周围形成了层次对比,突出的信息会向人的大脑传递这样一个信号:这里是重点,隐性地告诉受众这里是他们最应该关注的地方,使得受众能够更简单、更快速的处理创作者提供的信息。在数据故事中适当地突出创作者想要向受众强调的信息,可以让受众快速捕获到重要信息,在很大程度上提高数据故事的叙述效果。

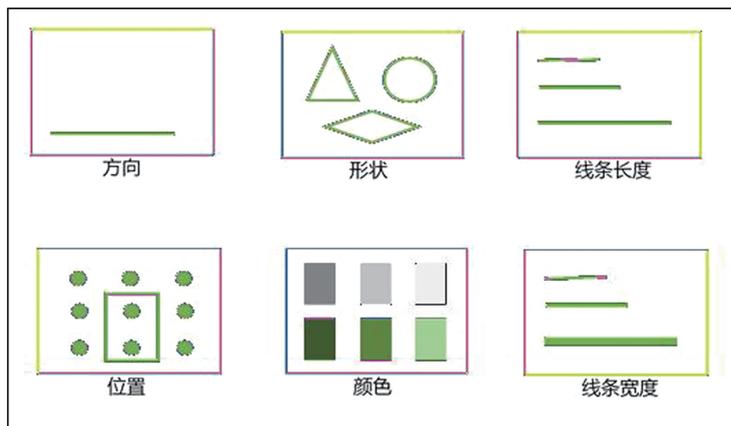


图5 常用视觉通道

如果图表设计过于杂乱,受众在阅读数据时很容易被分散注意力,受其他信息干扰。图表的设计应以简洁为核心,尽可能消除与中心思想无关的其他因素。具体策略包括:删除无效信息;利用对比不明显的颜色弱化干扰信息;使用较小的字体或形状弱化辅助信息等。

数据故事化描述的目的是吸引受众,激发受众的想象力。因此,不应在一张图表上呈现过多的信息,而应该在不同的版面依照故事流分开呈现不同的情节,激励有好奇心的受众逐步深入地探索数据并理解数据。

### 3.6 故事叙述

iRobot 数据科学总监 Bassa 在 Accelerate 会议上指出:“你必须引导受众对分析有一种直观的理解。从根本上说,这是一个沟通问题。如果你完成了数据整理与分析以及修改的所有工作,接下来你必须用数据进行沟通。”<sup>[50]</sup>在对数据图表进行优化之后,更重要的是将数据故事与受众进行叙述与交互,创作者应有效地综合数据故事,提供背景叙述,按情节引导受众。

数据故事化将数据转换为图形,饼图和折线图等图表使我们可以更直观地看到数据,但是单独的数据可视化具有局限性:只提供了一

一目了然的数据图表,缺乏解释事件发生原因的背景<sup>[51]</sup>。数据可视化只是让受众看到了数据,而不能激发受众采取进一步行动,这是因为受众只看到了一组冷冰冰的数据,知道这组数据表示的含义,却不知道数据的背景,无法从中产生共鸣。因此,作者需要为受众解释进行此次数据分析工作的原因,明确要解决的业务或组织问题,概述先前的相关工作<sup>[52]</sup>,使受众充分理解作者接下来要叙述的数据,带动受众的情感反应。

除了上述步骤之外,在实际的数据故事化项目中还可能涉及以下三个特殊活动。

(1)数据故事化的试验与预调研:为了达到更好的数据可视化目的,通常随机选取部分受众为测试样例,对即将采用故事化的故事模型和故事呈现方式进行测试和预调研,并根据试验与调研结果对数据故事化项目进行优化和调整。

(2)数据故事化的持续改进:与试验和预调

研不同的是,数据故事化的持续改进是根据最终故事化结果在全体受众中产生的效果和反馈结果改进数据故事模型及其呈现方式,进而动态改进数据故事化的效果。

(3)数据故事的作者与叙述者之间的分离:数据故事的创作者又涉及两类人才,一种是数据模型的设计者(数据故事的作者);另一种是数据故事的呈现者(数据故事的叙述者),分别负责从数据到故事模型和从数据模型到故事呈现的工作。在实际项目中,数据故事的作者和叙述者可以是同一个主体,也可以是不同的主体。当然,数据故事的作者或叙述者也可以是一个人,也可以是一个团队。

#### 4 数据故事的类型与呈现

数据故事化分为两个基本步骤:从数据到故事模型的转化(故事化建模)和从故事模型到故事的叙述(故事化呈现),如图6所示。

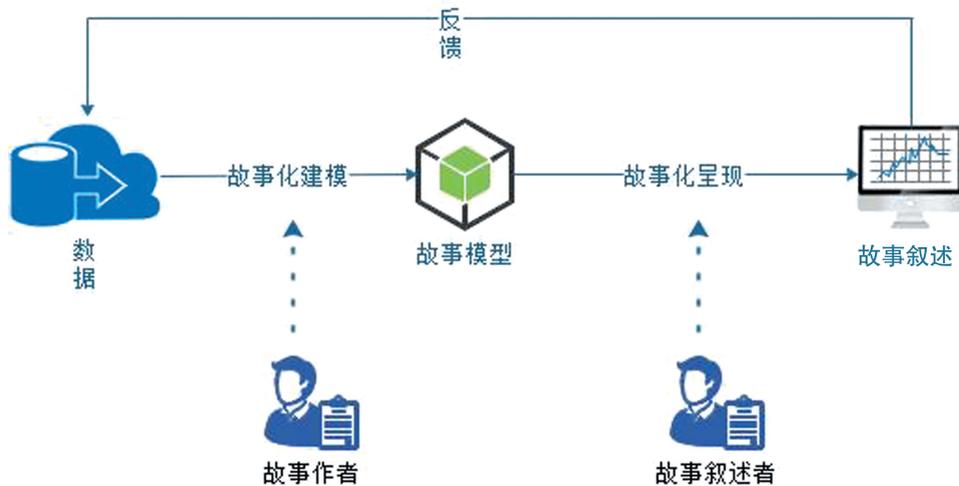


图6 数据的故事化描述及故事的呈现

(1)从数据到故事模型的转化是故事呈现的前提条件,同一个故事模型可以采用不同的呈现方式,如文章、图书、电影、海报、游戏、图片等,但表现效果可能不同。

(2)从故事模型到故事的叙述以及受众的

倾听过程也可以对数据建模产生反馈作用。数据故事模型的呈现过程往往可以发现数据故事化描述中存在的问题或漏洞,可为数据故事化提供反馈信息,从而进一步优化数据的故事化建模活动。

#### 4.1 数据故事的类型

数据的故事化描述结果是生成一个带有情节的数据故事, Davenport<sup>[32]</sup>、Tableau<sup>[33]</sup>和 Pros-

chool<sup>[34]</sup>分别从维度、场景和目的的角度对数据故事类型进行划分,如表4所示。

表4 数据故事的类型

视角	类型	描述
时间	过去式故事	呈现过去某一段时间数据的变动情况
	现在式故事	指出当前正在面临的情况
	未来式故事	预测未来发展
关注点	What 故事	说明数据呈现的客观事实
	Why 故事	分析现象背后的本质原因
	How 故事	提出解决问题的方法
深度	“CSI”故事	有针对性,短时间内就能讲述完
	“尤里卡”故事	故事很长、很重要、需要付出昂贵代价
方法	相关关系故事	揭示数据之间的相互关联性
	因果关系故事	揭示数据现象的原因和结果
场景	时间顺序	叙述方式及故事内容随着时间改变
	由大到小	用数据引导受众深入到某个聚焦的点
	由小到大	从小视角逆推到大视角
	突出对比	在数据集中强调不同之处,引出有力叙述
	探究交叉点	探究两条或多条数据曲线发生交叉的现象和原因
	因素划分	把主题划分为多个因素来剖析原因
目的	离群值	描绘出异常值,分析异常值背后隐藏的原因和原理
	说明性故事	生动形象地对数据现象进行描述
	解释性故事	用数据解释问题的原因及解决办法
	预测性故事	用已有数据预测接下来的趋势

(1) 基于维度。从时间维度划分,可以分为过去式故事——描述上周、上个月、上一季度或去年发生了什么的数据故事,大多数可视化分析故事也是这种类型的;现在式故事——分析人们或物体目前正在做什么,它实际上可能包括调查研究;未来式故事——使用预测分析,从过去获取数据来创建一个统计模型,然后用来预测未来。从关注焦点维度划分,可以分为“是什么(What)的故事”——就像报道故事一样,简单地叙述发生了什么;“为什么(Why)的故

事”——深入探究导致结果的潜在因素;如何解决问题(How)的故事——包含了“是什么(What)的故事”和“为什么(Why)的故事”,既有现状说明,又有原因解释,还对解决问题的办法进行了深入探究。从深度维度划分,可以分为:“CSI(Customized Small Investigation,小型专门调查)”故事——相对较小的特别调查,以找出为什么会发生次优的情况;“尤里卡(Eureka)”故事——对复杂问题的解决方案进行长时间的分析驱动探索。从方法维度划分,

分为相关关系故事——变量之间的关系同时上升或下降;因果关系故事——一个变量导致了另一个变量。

(2) 基于场景。针对不同的场景可以划分为七种类型的数据故事:随着时间的推移而变化——使用年来说明趋势;由大及小——设置上下文,以便受众能更好地了解特定类别中发生的事件;由小及大——描述受众关注的内容与大局的关系;突出对比——显示两个或更多主题之间的差异;探究交叉点——当一种类别超过另一种类别时突出重要的转变;因素划分——通过将主题划分为不同类型或类别来解释主题;离群值——显示事件的特别异常之处。

(3) 基于目的。从目的角度可以分为说明性数据故事、解释性数据故事和预测性数据故事。说明性数据故事——这样的故事数据丰富,并且只陈述与数据相关的事实,更多的是数据评论,然而事实需要以一种连贯的方式呈现出来,让人明白一个观点,决策者可以很容易地从中得出推论;解释性数据故事——探究业务问题,仔细研究细节,帮助决策者识别手头的真正问题,使受众通过数据理解问题背后的成因;预测性数据故事——将当前趋势的点连接起来,找到相同趋势的基本模式和原因,并预测事件的未来进程。

## 4.2 数据故事的呈现

数据故事通常包含两部分内容:叙述和图表。叙述是对故事背景的补充解释,图表是对数据分析结果的可视化呈现。数据故事的呈现是将数据故事以特定的形式展示给受众的过程,数据故事的呈现形式可分为口头叙述和演示报告两种。

(1) 口头叙述。叙述者直接将制作完成的数据故事讲给受众,受众以倾听的方式接受创作者口头叙述的数据故事。这种形式的优点是易于准备,在简单的会议室或者其他任何地方都能叙述,不需要复杂的场地和制作精良的可视化图表,只需要让叙述者充分理解数据的含

义并看懂自己制作的故事即可;缺点是对叙述者的要求较高,叙述者要理解故事并简单地表达出来,此外若叙述效果不好,受众则无法看到数据信息,对数据故事很难形成清楚的认知,只能听叙述者的陈述,容易受到叙述者主观因素的干扰。

(2) 报告演示。数据故事的展示过程中更多地采用报告演示的方式,传统的报告演示形式是叙述者向受众展示故事内容,同时对每一部分进行讲解和叙述,让受众既能“看到”也能“听到”。如今随着科学技术的发展以及受众对数据故事的需求增加,很多基于 Web 的数据故事只有创作者而没有叙述者,受众可以自行查看需要的信息。报告演示的形式可以分为静态演示和动态演示。其中,静态演示的数据故事形式通常分为文本形式和 PPT 形式两种。例如,The Global 绘制的 Starbucks 和 Dunkin 项目<sup>[53]</sup>将可视化和叙述有趣地结合在一起,逐步引导受众认识到“Starbucks 和 Dunkin 双方都在计划积极发展”。

动态演示的数据故事形式通常包括交互式图表和视频两种。交互式图表以一种更加灵活、智能、个性化、用户可操作的形式向受众叙述数据故事,受众可以根据自己的需求自由定义要查看的数据,交互式图表的创作者根据用户的需求制作并呈现相应的故事。这种形式的优点是可以增强受众的参与感,并且可以根据不同的受众进行个性化调整,使有不同需求的受众可以查看到自己所关心的信息,提高效率;缺点是对创作者要求较高,需要能够制作出合理美观的交互式图表。例如在 Baby Name Voyager 项目界面<sup>[50]</sup>,受众可以自由选择查看婴儿姓名与性别、年份之间的关系,点击姓名曲线还可以看到名字的由来、随时间推移该名字的人气变化以及关于该名字评论等。

视频形式包括短视频、电影、动画等,精心设计的、节奏良好的视频可以叙述一个有效的数据故事。优点是生动形象,受众可以看到可视化结果的动态变化,并且可以重复观看;缺点

是需要预先编制故事流的过渡,并对讲故事的节奏要有合理的掌握。例如,2006年Rosling在TED上的演讲“The best stats you’ve ever seen”<sup>[54]</sup>,使用了Gapminder制作图表动画,其中数据点的移动生动地揭示了不同国家的健康状况的改变。

## 5 数据故事化对数据科学的意义

数据故事化是数据科学的主要研究内容之一,也是数据科学区别于其他学科的重要特征。从流程看,数据科学项目包括数据化、数据加工、数据整齐化、探索性分析、数据分析与洞察、结果展现以及数据产品的提供等关键活动,而数据故事化是数据科学项目中结果呈现的主要手段之一。数据故事化对数据科学的意义主要体现在以下四个方面:

### 5.1 解决“最后一英里”问题

数据科学家精通编程、建模和数据清洗等数据处理业务,但还有一项关键技能经常被数据科学家们所忽视:清晰有效地传达研究结果的能力,即数据的故事化能力。数据的故事化是数据科学的“最后一英里”问题<sup>[5]</sup>。2016年LinkedIn的报告称,我们仍然生活在一个日益由数据驱动的世界,企业在积极招聘数据存储、数据检索和数据分析方面的专家<sup>[55]</sup>。但目前大部分招聘的重点都集中在数据准备和数据分析技能上,而不是数据科学的“最后一英里”技能<sup>[24]</sup>。

通过数据挖掘和数据分析等技术可以把数据价值挖掘出来,并把数据变成有价值的资源。但如果数据价值无法被受众有效吸收和利用,那么之前对数据所做的一切加工活动都失去了意义。根据Mayer和Anderson的研究,受众的视觉和听觉感官的刺激使他们对研究对象的理解提高了74%<sup>[41]</sup>。数据科学的重点是在向受众传达数据分析结果时进行沟通或说服<sup>[56]</sup>,将数据中提取的价值传递出去,数据的故事化是实现这一目的的极好手段。利用数据故事化,受

众能更深刻地认识数据、理解数据、记住数据并根据数据采取行动,充分发挥数据的价值。

### 5.2 从数据中获取洞察

数据科学的目标是提供可靠的基于数据的信息,并将信息转化为人们的宝贵见解,但如果缺乏适当的信息沟通,数据科学家就无法向受众提供价值<sup>[5]</sup>。使用图表和故事可以显示数字无法清晰描绘的要素和数据之间的关系,能够使受众深入理解数据传达的含义,获取数据洞察。

数据可视化有助于我们理解数据,但它只是让受众看到数据、认识数据,很难让受众体会数据、记住数据。数据故事化结合数据、可视化和叙述三个要素。数据可视化是叙述数据故事的最佳工具,因为它提供了可信度并且能对故事中的要点给予重视<sup>[40]</sup>。数据可视化的效果取决于它们所揭示的洞察以及受众记住这些洞察的时间<sup>[57]</sup>,而扣人心弦的故事化叙述是让受众记住这些数据洞察的高效方式,因此将数据故事化与数据的可视化相结合,更有助于受众获取数据洞察。

### 5.3 将数据洞察转换为数据行动

数据科学家的动机是让受众根据从数据中提取的见解采取行动,这很大程度上取决于数据科学家向最终用户或决策者提供数据的能力。遵循“认识问题—解决问题”的中心思想,数据的故事化就是在最合适的时间以适当的格式向受众呈现正确的信息<sup>[34]</sup>,并让受众对这些信息做出反应:过程分析、业务决策或提出讨论。

数据科学家最基本、最普遍的技能是编写代码的能力,但数据科学家每天都需要向利益相关者简单易懂地解释复杂模型的结果<sup>[28]</sup>,因此对数据科学家而言,更重要的是能用所有利益相关者都能理解的语言进行交流,无论是口头上的、视觉上的、还是口头与视觉相结合的方式<sup>[42]</sup>,数据科学家需要从故事叙述者那里吸取经验,用简单有趣的故事连接起数据和业务之间的断层,将数据分析和建模的结果以可理解

的格式传递给正确的受众。

#### 5.4 数据新闻及数据产品开发

数据故事化在依赖数据进行决策的领域能发挥更好作用,目前,数据故事化主要应用在新闻业和商业领域。数据新闻是当代新闻的一个方面,它使用数据以最佳方式叙述故事,且是开放的、可访问的和启发性的<sup>[58]</sup>,其中除了传统的新闻方法之外,还使用数据分析、编程和叙述可视化等技术来创建数据故事<sup>[59]</sup>。作为数字新闻的一项特色活动,数据的故事化正迅速得到重视,并被大小媒体机构广泛采用<sup>[60]</sup>。目前数据新闻主要用于政治、金融和事实调查等领域。

## 6 结论

目前,数据的故事化已成为获取并分享数据洞察的最佳方式之一,是实现从数据感知到数据认知转化的主要手段。自从印刷术、电报、广播、电影和计算机发明以来,这些技术不断地为我们提供新的、越来越复杂的讲故事的手段,尤其是计算机的出现,使娱乐、艺术和技术以一种有趣的方式融合在一起,推动了数据故事化的研究进展。

现阶段有关数据故事化研究的主要特点是:①研究多集中在国外,国内对该领域的学术研究几乎没有;②在国外的研究成果中,主要以博文等非正式发表的文章为主,在正式出版物

上公开发表的学术论文较少;③理论研究滞后于实践应用,且尚未有系统深入的研究成果;④实践应用过度雷同,缺少新的突破,多为理念层面上的修正和优化,缺乏具体的实现技术;⑤数据故事化工具多依托可视化工具,专门的数据故事化工具较少。

数据故事化是一个值得深入探索的新课题,未来研究可从以下几个方面进行:①完善数据故事化的理论体系,即包括数据故事化的理念、理论、方法、技术、工具、实践在内的一整套完整的知识体系;②研究数据故事化的评价方法,引入过程成熟度评估方法,提升数据故事化的信度和效度,并对具体的数据故事项目进行优化与改进;③加强多学科交叉研究,凝聚来自数据科学、认知科学、可视分析学、信息图形学、自然语言理解、人机交互等不同学科领域的研究成果,进一步拓宽数据故事化的研究视角和理论基础;④探索数据故事化的新算法和新模型,模型和算法是数据故事化处理的核心,但是目前对面向数据故事化的新算法或新模型的研究并不多,亟待研发出面向不同应用场景的核心算法或模型及其自动调整参数方法;⑤研发面向数据故事化的专门工具,目前,数据故事化的软件工具不多且功能不够成熟,主要作为数据可视化软件工具的一个辅助功能的形式存在,缺少专用于数据故事化的软件工具,尤其是开源工具。

## 参考文献

- [1] Zimmer J. Making it stick: tell stories[EB/OL]. [2019-03-25]. <https://mannerofspeaking.org/2009/10/13/making-it-stick-tell-stories>.
- [2] 朝乐门. 数据科学理论与实践[M]. 北京:清华大学出版社,2017.(Chao Lemen. Data science theory and practice[M]. Beijing: Tsinghua University Press, 2017.)
- [3] Munn K. Made to stick for the social sector: the quest for messages that resonate[EB/OL]. [2019-03-25]. <https://centers.fuqua.duke.edu/case/2015/11/13/made-to-stick-for-the-social-sector-the-quest-for-messages-that-resonate/>.
- [4] Kosara R, Mackinlay J. Storytelling: the next step for visualization[J]. Computer, 2013, 46(5): 44-50.
- [5] Roma I. Using data to deliver a message: storytelling with data[EB/OL]. [2019-02-20]. <https://medium>.

- com/8wires/using-data-to-deliver-a-message-storytelling-with-data-b490b7b2c9bd.
- [ 6 ] Welbourne T M. Data-driven storytelling: the missing link in HR data analytics[J]. *Employment Relations Today*, 2015, 41(4): 27-33.
- [ 7 ] Drew S E, Duncan R E, Sawyer S M. Visual storytelling: a beneficial but challenging method for health research with young people[J]. *Qualitative Health Research*, 2010, 20(12): 1677-1688.
- [ 8 ] Underwood J. Learning the art of analytical storytelling[EB/OL]. [2019-04-03]. [https://www.sas.com/id\\_id/insights/articles/business-intelligence/learning-the-art-of-analytical-storytelling.html](https://www.sas.com/id_id/insights/articles/business-intelligence/learning-the-art-of-analytical-storytelling.html).
- [ 9 ] Thue D, Bulitko V, Spetch M, et al. Interactive storytelling: a player modelling approach[C]//AIIDE. 2007: 43-48.
- [ 10 ] Turner S R. The creative process: a computer model of storytelling and creativity[M]. New York: Psychology Press, 2014: 4-20.
- [ 11 ] Knaflig C N. Storytelling with data: a data visualization guide for business professionals[M]. John Wiley & Sons, 2015:96-99.
- [ 12 ] Pretty C. Use data and analytics to tell a story[EB/OL]. [2019-02-20]. <https://www.gartner.com/smarter-withgartner/use-data-and-analytics-to-tell-a-story/>.
- [ 13 ] Mayer R E, Moreno R. Nine ways to reduce cognitive load in multimedia learning[J]. *Educational Psychologist*, 2003, 38(1):43-52.
- [ 14 ] Yuvaraj M. Infographics: tools for designing, visualizing data and storytelling in libraries[J]. *Library Hi Tech News*, 2017, 34(5):6-9.
- [ 15 ] The Presentation Company. The ultimate guide: data visualization vs. storytelling with data: what's the difference?[EB/OL]. [2019-05-24]. <https://blog.presentation-company.com/data-visualization-vs.-data-storytelling-whats-the-difference>.
- [ 16 ] Gotz D, Zhen W. Behavior-driven visualization recommendation[C]// International Conference on Intelligent User Interfaces. 2009.
- [ 17 ] Samuel A. How content marketers can tell better stories with data[EB/OL]. [2019-07-07].<https://hbr.org/2015/12/how-content-marketers-can-tell-better-stories-with-data>.
- [ 18 ] Weber W, Engebretsen M, Kennedy H. Data stories: rethinking journalistic storytelling in the context of data journalism[J]. *Studies in Communication Sciences*, 2018, 18(1):191-206.
- [ 19 ] Google. Data journalism in 2017: the current state and challenges facing the field today[EB/OL]. [2019-05-25]. <https://news.google.com/assets/docs/data-journalism-in-2017.pdf>.
- [ 20 ] Nichols N. Data storytelling: the interface of the past, present and future[EB/OL]. [2019-05-25]. <https://narrativescience.com/blog/data-storytelling>.
- [ 21 ] Harrold L. Let your data tell a story[EB/OL]. [2019-05-25]. <https://www.ttec.com/articles/let-your-data-tell-story>.
- [ 22 ] French K. Why data visualization + data storytelling is marketing gold[EB/OL]. [2019-05-25]. <https://www.columnfivemedia.com/data-storytelling-brands-data-visualization>.
- [ 23 ] Davenport T. Why data storytelling is so important—and why we're so bad at it[EB/OL]. [2019-05-26]. <https://www2.deloitte.com/insights/us/en/topics/analytics/data-driven-storytelling.html>.

- 
- [24] Dykes B. Data storytelling: the essential data science skill everyone needs[EB/OL]. [2019-05-26]. <https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/#6372f64152ad>.
- [25] Segel E, Heer J. Narrative visualization: telling stories with data[J]. IEEE Transactions on Visualization and Computer Graphics, 2011, 16(6):1139-1148.
- [26] Analytics Vidhya Content Team. The art of story telling in data science and how to create data stories?[EB/OL]. [2019-05-27]. <https://www.analyticsvidhya.com/blog/2017/10/art-story-telling-data-science/>.
- [27] Wikipedia. Plot (narrative)[EB/OL]. [2019-05-27]. [https://en.wikipedia.org/wiki/Plot\\_\(narrative\)](https://en.wikipedia.org/wiki/Plot_(narrative)).
- [28] Zawadzki J. Storytelling for data scientists: turn data into stories to persuade your audience[EB/OL]. [2019-05-27]. <https://towardsdatascience.com/storytelling-for-data-scientists-317e2723aa31>.
- [29] Chibana N. Mastering data storytelling: 5 steps to create persuasive charts and graphs[EB/OL]. [2019-05-27]. <https://www.crazyegg.com/blog/data-storytelling-5-steps-charts/>.
- [30] Taborga D. 6 steps to data storytelling[EB/OL]. [2019-05-27]. <https://blogs.iadb.org/conocimiento-abierto/en/6-steps-to-data-storytelling>.
- [31] Maptive. The 101 guide to telling a compelling story with data[EB/OL]. [2019-05-27]. <https://www.maptive.com/101-guide-telling-compelling-story-data/>.
- [32] Davenport T. 10 Kinds of stories to tell with data[EB/OL]. [2019-07-07]. <https://hbr.org/2014/05/10-kinds-of-stories-to-tell-with-data>.
- [33] Tableau. Best practices for telling great stories[EB/OL]. [2019-05-28]. [https://onlinehelp.tableau.com/current/pro/desktop/en-us/story\\_best\\_practices.htm](https://onlinehelp.tableau.com/current/pro/desktop/en-us/story_best_practices.htm).
- [34] Proschool. Data storytelling: a critical skill for a data scientist[EB/OL]. [2019-05-28]. <https://www.proschoolonline.com/blog/data-storytelling-critical-skill-data-scientist>.
- [35] Data visualization 101: how to design charts and graphs[EB/OL]. [2019-05-28]. [https://cdn2.hubspot.net/hub/53/file-863940581-pdf/Data\\_Visualization\\_101\\_How\\_to\\_Design\\_Charts\\_and\\_Graphs.pdf](https://cdn2.hubspot.net/hub/53/file-863940581-pdf/Data_Visualization_101_How_to_Design_Charts_and_Graphs.pdf).
- [36] Evergreen S, Emery A. Data visualization checklist[EB/OL]. [2019-05-29]. [https://datavizchecklist.stephanieevergreen.com/assets/DataVizChecklist\\_Feb2018.pdf](https://datavizchecklist.stephanieevergreen.com/assets/DataVizChecklist_Feb2018.pdf).
- [37] Tableau. The evolution of data storytelling[EB/OL]. [2019-05-29]. [https://narrativescience.com/wp-content/uploads/2018/12/Whitepaper\\_Tableau\\_The-Evolution-of-Data-Storytelling.pdf](https://narrativescience.com/wp-content/uploads/2018/12/Whitepaper_Tableau_The-Evolution-of-Data-Storytelling.pdf).
- [38] Veel K. Make data sing: the automation of storytelling[J]. Big Data & Society, 2018(5):1-8.
- [39] Michetti N. How data will influence the future of storytelling[EB/OL]. [2019-05-29]. <https://powerbi.microsoft.com/fr-fr/blog/how-will-data-influence-the-future-of-storytelling/?cdn=disable>.
- [40] Dietz K. Data storytelling: what it is, what it's not, how to do it[EB/OL]. [2019-05-29]. <https://www.just-storyit.com/inside-story-blog/2017/11/30/data-storytelling-what-it-is-how-to-do-it-how-to-think-about-it>.
- [41] Forbes. Video in the C-suite: executives embrace the non-text web[EB/OL]. [2019-05-29]. [https://images.forbes.com/forbesinsights/StudyPDFs/Video\\_in\\_the\\_CSuite.pdf](https://images.forbes.com/forbesinsights/StudyPDFs/Video_in_the_CSuite.pdf).
- [42] Recruiting M. Data scientist: the sexiest job of the 21st century[J]. Harvard Business Review, 2012, 90(10):70-76.
- [43] Dataquest. Building a data science portfolio: storytelling with data[EB/OL]. [2019-05-29]. <https://www>.
-

- dataquest.io/blog/data-science-portfolio-project/.
- [44] Watson H. Data visualization, data interpreters, and storytelling[J]. *Business Intelligence Journal*, 2017(22): 5-10.
- [45] Gartner J. Hype cycle for data science, 2016[EB/OL]. [2019-04-20]. <https://www.gartner.com/doc/3388917/hype-cycle-data-science->.
- [46] Mannon N. Persuasive storytelling with data visualization[EB/OL]. [2019-04-20]. <https://www.blastam.com/blog/persuasive-storytelling-with-data-visualization>.
- [47] Anscombe F J. Graphs in statistical analysis [J]. *The American Statistician*, 1973, 27(1): 17-21.
- [48] Doran G T. There's a smart way to write management's goals and objectives[J]. *Management Review*, 1981, 70(11): 35-36.
- [49] Stikeleather J. How to tell a story with data[EB/OL]. [2019-03-01]. <https://hbr.org/2013/04/how-to-tell-a-story-with-data>.
- [50] TDWI. TDWI accelerate seattle October 16-18[EB/OL]. [2019-03-10]. <https://tdwi.org/events/accelerate/boston/information/agenda.aspx>.
- [51] Nugit. What is data storytelling?[EB/OL]. [2019-03-10]. <https://www.nugit.co/what-is-data-storytelling>.
- [52] McNulty K. Beware of 'storytelling' in data and analytics[EB/OL]. [2019-03-13]. <https://towardsdatascience.com/beware-of-storytelling-with-data-1710fea554b0>.
- [53] Chang A, Carroll M. Split country: Dunkin'vs. Starbucks[EB/OL]. [2019-03-06]. [http://archive.boston.com/yourtown/specials/starbucks\\_vs\\_dunkin\\_donuts/](http://archive.boston.com/yourtown/specials/starbucks_vs_dunkin_donuts/).
- [54] Rosling H. The best stats you've ever seen[EB/OL]. [2019-03-07]. [https://www.ted.com/talks/hans\\_rosling\\_shows\\_the\\_best\\_stats\\_you\\_ve\\_ever\\_seen](https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen).
- [55] Murthy S. The 25 skills that can get you hired in 2016[EB/OL]. [2019-03-09]. <https://blog.linkedin.com/2016/01/12/the-25-skills-that-can-get-you-hired-in-2016>.
- [56] Swoyer S. Enable deeper understanding with great data storytelling[EB/OL]. [2019-03-16]. <https://tdwi.org/articles/2017/03/09/enable-deeper-understanding-with-great-data-storytelling.aspx>.
- [57] Ryan L. Data visualization and data storytelling: a visual revolution[EB/OL]. [2019-03-16]. <http://www.dbta.com/BigDataQuarterly/Articles/Data-Visualization-and-Data-Storytelling-A-Visual-Revolution-124077.aspx>.
- [58] Rogers S. Data journalism matters more now than ever before[EB/OL]. [2019-03-20]. <https://simonrogers.net/2016/03/07/data-journalism-matters-more-now-than-ever-before/>.
- [59] Appelgren E, Nygren G. Data journalism in Sweden[J]. *Digital Journalism*, 2014, 2(3): 394-405.
- [60] Ojo A, Heravi B. Patterns in award winning data storytelling[J]. *Digital Journalism*, 2018(6): 693-718.

**朝乐门** 中国人民大学信息资源管理学院副教授,数据工程与知识工程教育部重点实验室(中国人民大学)研究员,博士生导师。北京 100872。

**张晨** 中国人民大学信息资源管理学院、数据工程与知识工程教育部重点实验室硕士研究生。北京 100872。

(收稿日期:2019-04-17;修回日期:2019-06-02)