

算法创作： 大数据时代电视节目创作的新模式

——以《纸牌屋》为例

✍ 王朋进

随着奈飞公司(Netflix)制作的《纸牌屋》在网上热播,大数据对于电视节目创作的神奇作用越来越引起业界的重视。一些人甚至认为这是“算法节目制作”(Algorithmic Programming)时代来临的标志,将颠覆传统的电视节目创作模式,甚至艺术家的饭碗也将因此被夺走。拥有大数据优势的互联网企业跃跃欲试:阿里巴巴与东方卫视合作了综艺节目《女神的新衣》、腾讯推出了辩论类新闻节目《事实说》、腾讯与唯众合作制作了网络节目《你正常吗?》、乐视公司成立了乐视影业,盛大文学也成立编剧公司,依托“大数据”创作电视节目。

面对这一“突如其来”的新事物,电视界多少有些愕然。乐观者视之为一个电视节目创作新时代的到来,而悲观者却视之为艺术创作的“末日”(Apocalypse),是技术对艺术的亵渎。至于究竟应该如何掌握这一新手段,则几乎没有像样的回应。那么,《纸牌屋》的大数据成功故事究竟是如何演绎的?大数据到底会对电视节目创作带来怎样的影响?未来的电视节目创作是否会被大数据

“绑架”?“算法”技术文明到底会对电视节目创作艺术产生怎样的冲击?本文将对这些问题进行探讨和思考。

一、技术神话还是营销噱头——《纸牌屋》创作经验的再认识

提到大数据用于电视节目创作,《纸牌屋》(House of Cards)无疑是最为耀眼的一个案例。的确,这部剧获得了火热的市场反应,但不同于其他热播的电视剧,人们谈论这部剧时其实有两大话题主线:一是剧作本身(故事情节、人物性格等),二是它的大数据技术特点——甚至关于大数据技术的热衷要超过对剧作本身。这样一部具有神奇技术色彩的作品,业界和学界对如下问题不可能不感兴趣:这部以大数据技术制作著称的电视剧究竟是如何诞生的?制作者是如何将大数据融进电视剧制作的?大数据在这部剧作的成功中到底发挥了怎样的作用?然而遗憾的是,诸多介绍的文章并没有对这部剧作的制作神奇之处给出详尽分析,仅是

停留在对技术神化的感叹上。为此,我们有必要首先来重新认识一下《纸牌屋》的创作经验。

《纸牌屋》是美国著名的在线影片租赁商奈飞公司(Netflix)2013年推出的一档39集的政治题材电视剧。这部电视剧由詹姆斯·弗雷、大卫·芬奇等执导,鲍尔·威利蒙改编,凯文·史派西、罗宾·怀特、迈克尔·凯利、拉斯·米科尔森、凯特·玛拉、克里斯汀·康诺利等主演。该剧讲述一个冷血无情的美国国会议员及与他同样野心勃勃的妻子在华盛顿白宫中运作权力的故事。《纸牌屋》第一季于2013年2月1日在Netflix网站上全球同步首播,第二季于2014年2月14日播出,第三季于2015年2月27日播出,第四季于2016年3月4日在美国播出,第五季也已确定续拍。在全美国两亿多互联网用户中,该剧第一季的平均收视人口达到960万,第二季的第一集收视人口就达到了1340万,最高时有超过3000万用户收看。

那么这部电视剧是怎么创作出来的呢?

首先,该剧的创作动因来自大数据分析。奈飞公司通过其标签数据和推荐系统对其3000多万用户的行为数据分析后发现,由演员凯文·史派西(Kevin Spacey)和导演大卫·芬奇(David Fincher)联手拍摄的政治题材的连续剧最受用户欢迎。Netflix数据库包含3300万用户的收视选择、400万条评论、300万次主题搜索。这些大数据显示,20年前的旧剧《纸牌屋》仍然得到观众的喜爱,而且这些观众也几乎和网站上大卫·芬奇、凯文·史派西的粉丝圈重合。这一数据结果给了奈飞公司投资创作的勇气和决心。

其次,剧本是原来就有的,而非原创。剧本是一剧之本,从原创性的角度来说,剧本的原创性具有决定性。影视史上不乏好故事被反复重新拍摄的例子,但翻拍作品毕竟脱不了原有的故事框架,因而也算不上彻底的原创。就这一点而论,《纸牌屋》在剧本的创作上其实并没有体现任何大数据的作用。如果一定要找出和大数据的关系的话,我们或许可以这么看:大数据让创作者重新发现了这个剧本——原来以为它早已不在观众喜好之中,却没想到观众对它念念不忘。当然,因为这次要给美国观众看,所以剧本还是要做一些改编。

再次,主创人员的选择也跟大数据有关。大数据发现了有趣的关联:奈飞的数据提供三个结论:1、有很多用户通过社交网络完整收看了由大卫·芬奇导演的影片。2、英国版的《纸牌屋》依旧受到欢迎。3、那些看过英国版《纸牌屋》的人还观看了凯文·史派西主演和(或)大卫·芬奇执导的影片。这样,导演和主演的人选就可以确定下来。其他重要的人选也基本可以通过奈飞的大数据分析逐一找到。

这几乎为该剧的成功提供了全部的理由。

然而,我们还是忍不住要提出一些思考:究竟哪些算得上大数据的贡献?大数据究竟是否具有预测性?大数据把观众的心思都洞察了吗?笔者认为这个案例还难以给我们一个充分肯定的答案。因为上述三点其实即便没有大数据,在我们传统的电视剧创作中也不乏成功案例。好的剧本之所以会被反复翻拍,是因为创作者心里知道它有群众基础。至于导演和演员的票房号召力,这些都已经过去的市场实践中有了很好的注脚。换句话说,大数据至少到此为止并没有提供足够的证据证明其在这部剧的创作中具有非凡的不可替代的创作能力。而且这里还有一个关键的问题没有回答:奈飞投资一亿美元的高昂代价(单集成本是同类型作品的两倍)拍摄《纸牌屋》第一季,在剧作本身的商业上是成功的吗?因此也就难怪有人质疑大数据只是营销的噱头。

但它确实在一个极为重要的问题上给予创作者巨大的支持,那就是制作决策。大数据可能并没有在具体的节目创作过程中表现出多大的作用,但在要不要拍以及拍什么样的剧作方面给创作者提供了重要的决策依据。这个决策依据的分量还是相当重的。因此,我们不能简单地认为大数据用于电视节目创作只是一个营销噱头,它确实在发挥作用。

二、迎合还是引领——算法技术与电视节目创作的本质

但很显然,我们也不宜把大数据的作用过于夸大,因为迄今为止的大数据智能跟踪技术相比于人的艺术创作存在一个致命的缺陷。严格来说,大数据技术提供的不是创造性的、开掘性的,它只是跟随性的、迎合性的。或者直白地说,大数据技

术只能迎合用户的口味，并不能引领用户的品味。这是它与人的艺术创作最本质的区别。

我们先来了解一下大数据的特点。按照舍恩伯格的理解，大数据（Big Data）指的是不用随机分析法（抽样调查）这样的捷径，而采用对所有数据进行分析处理的一种数据分析模式。因此大数据具有五大特点：Volume（巨量）、Velocity（高速）、Variety（多样）、Value（低价值密度）、Veracity（真实性）。传统的随机抽样数据分析和现代大数据分析技术相比有相似性，更有差异性。比如：两者都是对行为结果的分析（是跟踪分析而不是预测性分析），只是在技术模式上不同而已。过去由于数据收集和分析方面的限制，着重于强调数据分析的效率，随机抽样法比较能够满足这一需求。但这并不意味着随机抽样是完美的方法，因为随机抽样毕竟还存在可靠性的概率问题。

以电视收视率分析为例，传统分析方法是随机抽样的，样本量几千（有的城市少的只有几百样本）而已，数据量通常不过数十万，在可容忍的精确度（误差率）范围之内，收视率调查结果是可以接受的。在互联网领域，我们在网络上的每一个动作都会留下电子痕迹，这个电子痕迹就是被网络平台自动记录的电子数据，所有用户的电子数据汇总在一起就是所谓的“大数据”。因此，大数据是全样本分析，数据量动辄上亿。在传统技术条件下，这样巨量的数据收集和分析需要耗费资源是惊人的，但在互联网数据传播技术普及的今天，对于那些大型网络平台而言，却可以很方便快捷地完成。

但电子痕迹并不都是有用数据，其中有很多只是垃圾数据。如何从一大堆乱七八糟的数据中提取出有用的数据，这需要网络平台优化算法结构。因此，这正是大数据用于电视节目创作被称为算法创作的原因，也是各大网络平台竞争的焦点。从搜索引擎到兴趣跟踪再到人工智能，其技术核心都是算法问题。也就是说，谁的算法更优化，谁就更具技术优势。我们可以想象一下，“双十一”一天24小时阿里巴巴网络平台所承受的数据量之大，如果没有强大的算法优势，平台可能早就瘫痪了。

电视节目创作的本质是什么？是运用电视技术

手段创作群众喜闻乐见的作品。按照我们的主流价值观，这里的喜闻乐见既包括了市场效益的丰收，也包括社会价值（包括艺术价值）层面的认可。因此，它不是单纯的口味迎合，还要实现价值引领。电视节目制作最难的莫过于了解观众真正的兴趣喜好。观众对什么题材感兴趣？对什么节目样式感兴趣？观众的消费心理周期是怎样的？这些都像谜一样困扰着每一个创作者。就大数据技术应用目前所具备的能力而言，它在寻找、迎合观众趣味方面能够发挥明显的优势（这对于电视节目创作当然也是很重要的），但在艺术品位提升方面几乎无能为力。它对于已然的事物有细致的洞察，但对于未然的事物几乎不具备可靠的预测。

深耕大数据的平台企业都知道数据结构化的重要性，只有结构化的数据才是真正有价值的数，否则一大堆数据只能是无法使用的垃圾。所以，优化算法进而获得优质的结构化大数据成为平台在这一领域竞争的焦点。标签化是其中一个重要的动作，给电视节目配以尽可能多的标签，然后在巨量的点击接触数据中分析各个标签的数据矩阵，得出优先次序的结论，这就是大数据提供给创作者的决策依据。

为了优化自己的算法，奈飞公司连续几年专门成立了一个百万美元的奖项来吸引全世界优秀的数学家、电脑专家、人工智能专家来帮助它。因此，奈飞公司可以从已有的数百个标签的数千万（甚至数亿）个使用数据的结构化分析中比较准确地了解它的用户具体的喜好。这也是奈飞对自己的大数据分析颇为自信的原因——否则它怎么会敢于投资一亿美元来翻拍《纸牌屋》？但一如前面的分析所呈现出来的，奈飞的大数据尽管丰富而精致，但也只能起到迎合的作用，它无法实现引领。因此，我们也可以说：大数据用于电视节目创作是一个重要的技术辅助，它和艺术创作并不冲突——它取代不了艺术创作。

三、红利的正确打开方式——大数据如何更好地服务节目创作

先进的技术总是具有一定的红利，谁先掌握谁就获得享受红利的优先权。大数据对于消费者

口味的准确跟踪对于创作者来说是个极大的诱惑,怎样才能享受这样的技术红利是一个具有现实性的问题。许多电视台也都有自己的网络平台,每天也在产生大量的数据,这样的数据是否就是我们向往的大数据?对于纯粹的制作机构而言,大数据又从何而来?而且特别关键的是:大数据要花多少钱?毕竟,奈飞投资一亿美元的《纸牌屋》在剧作本身的商业上是否成功是很受质疑的。或者说,我们能得到优质价廉的大数据服务吗?如果不能,它还只能是神坛上的祭品,只能看不能吃。

首先,“很多数据”并不等于“大数据”。大数据确实表现为数据量的“大”,但不是一般的大,而是巨量的大。而且大数据的质量也跟数量规模有关,规模越大质量越高。虽然没有具体的标准,但我们可以和普通的、传统的调查方式产生的数据量进行比较。普通一次调查(随机抽样那种)的数据量大约数万,至多数十万。而大数据的数据量至少在这个基础上增加几个数量级——数百万、数千万,乃至数亿。比如,奈飞公司基于3300万用户使用痕迹的大数据就可以评估节目创作的可行性,但若它只是330万用户基础上的使用数据恐怕就不敢放心决策。这样一来,只有少数网络巨头才真正拥有这个级别的大数据。

其次,非结构化的大数据是初级产品,要想适合使用还需要进行结构化、半结构化处理。大数据真的成为资产、商品,必须进行深加工——针对不同的需求,提炼出所需的大数据,建立相应的数据库。随便给你一堆数据,那是一团乱麻,是没有什么用的,你需要的是为你提炼好的专门的数据库服务。比如:我想知道爱看足球的男生更喜欢哪种类型的电视剧,爱看足球的女生喜欢哪种服饰?这就需要技术人员在数据存储的技术设计中进行优化的计算处理。这也是包括奈飞在内的致力于大数据资产运营的平台正在努力做的事情。

这样一看,大数据服务于节目创作的门槛就很高了。一是能提供大数据服务的机构不多。目前,大型互联网平台(比如:国外的Facebook、Twitter、Google,国内的腾讯、阿里巴巴、百度等)通常拥有巨量的用户大数据信息。这些机构注册用户动辄数亿,每天产生惊人数量的大数据。二

是价格不可能很低。因为涉及复杂的技术条件,想要得到针对性的、结构化处理的实用大数据服务必然要付出昂贵的代价。奈飞投资制作《纸牌屋》因为使用的是自己拥有的大数据,所以尽管投资一亿美元,里面并不包括大数据本身的价格。如果把大数据的价格也计算上,恐怕电视剧的成本还要更加昂贵。

这里面就产生了一个很现实的大问题:我们究竟愿意为大数据服务付出多大代价?如果大数据能够给我们准确评估出市场产出,这个投入就是理性的;假如不能做到,那么这种投资就是冒险,也就没有普遍意义。《纸牌屋》的案例价值受到的最大怀疑正是这一点。因为奈飞并不是靠播出广告获益的商业模式,所以它需要通过增加用户(订阅费)来赚钱,以7.99美元的单集订阅费来计算,一亿美元需要增加1000多万用户才打平。即便以美国当红电视节目的商业模式来计算,单集高达400—600万美元的成本(美国当红电视剧单集价格大约200—300万美元)也是风险极大的,而且这还没有算上大数据服务的成本。因此,我们只有寄希望于大数据服务价格的下降。

对于大数据平台企业而言,数据固然是资产,但若不能产生效益,其实算不上真正意义上的资产,反而可能是负担。多方面开发大数据的商业价值,多吸引用户当然成为首要考虑的目标。一种方式是自主开发,像奈飞公司那样。但奈飞公司有做电视节目的优势:这本来就是它的经营主业。像阿里巴巴这样的平台企业就未必合适。另一种方式是合作开发,我出大数据,你来搞创作,风险共担,利益共享。这样,平台企业可以把数据“卖”出去,而节目开发机构又可以避免为购买大数据支付昂贵的价格。在这一合作模式中,平台企业应该定位于数据服务,而不是创作决策,也就是说,要明确大数据的服务角色,不能本末倒置。这样才能形成优势互补,技术红利也才可能成为现实的可分享的美味。

(作者系中国农业大学人文与发展学院媒体传播系副教授、硕士研究生导师)

(责编 一申)