

The Construction of Video Content Evaluation Model

Based on the Big Data of Omnimedia

基于全媒体大数据的

视频内容评估模型建构

吴殿义 著

清华大学出版社

The Construction of Video Content Evaluation Model

Based on the Big Data of Omnimedia

基于全媒体大数据的 视频内容评估模型建构

吴殿义 著

清华大学出版社
北京

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

基于全媒体大数据的视频内容评估模型建构 / 吴殿义著. —北京: 清华大学出版社, 2020.4

ISBN 978-7-302-53368-9

I. ①基… II. ①吴… III. ①视频编辑软件—数据处理—研究 IV. ①TP317.53

中国版本图书馆 CIP 数据核字(2019)第 168342 号

责任编辑: 纪海虹

封面设计: 甘 玮

责任校对: 王凤芝

责任印制: 杨 艳

出版发行: 清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址: 北京清华大学学研大厦 A 座 邮 编: 100084

社 总 机: 010-62770175 邮 购: 010-62786544

投稿与读者服务: 010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈: 010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者: 三河市吉祥印务有限公司

经 销: 全国新华书店

开 本: 165mm×240mm 印 张: 9.75 字 数: 159 千字

版 次: 2020 年 4 月第 1 版 印 次: 2020 年 4 月第 1 次印刷

定 价: 58.00 元

产品编号: 078208-01

目 录

绪论	1
第一节 研究缘起与背景	1
一、内容市场的变化	1
二、学界的探索	2
三、长期的追踪性研究为本研究提供了支撑	2
第二节 前人研究与文献综述	3
一、内容评估方面的研究	3
二、大数据的相关研究	9
三、全媒体的相关研究	15
四、内容银行的前人研究	19
第三节 视频内容评估产品的发展及现状	20
一、内容评估产品发展的三个阶段	21
二、内容评估产品的三重问题	29
第四节 研究方法及框架	30
一、文献研究	32
二、定性分析	32
第一章 内容评估体系建构的基础	34
第一节 市场基础：大视频产业对内容评估提出需求	34
一、视频内容产业进入多元竞争格局,原有生存法则发生变化, 需要评估体系支撑	34
二、缺乏评估体系的内容交易模式不能满足视频内容产业的 需求	39

第二节	技术基础：日益成熟的大数据产业为评估体系 提供了现实可能	44
一、	数字技术始终是内容评估发展的驱动力	45
二、	大数据产业的发展日趋成熟	47
第三节	本章小结	54
第二章	全媒体大数据内容评估体系的模型建构	56
第一节	基于全媒体大数据的内容评估模型的原则	56
一、	满足内容产业全流程评估的需求	56
二、	与内容生产相关各要素的价值评估要计入内容评估 体系中	58
三、	利用大数据,结合主观经验评估	60
四、	充分挖掘数据价值,实现定量与定性结合	61
第二节	基于全媒体大数据的视频内容评估模型建构	62
一、	内容评估体系由 5 个模块组成	62
二、	全媒体收视模块	64
三、	全媒体社交舆情模块	65
四、	全媒体传播模块	67
五、	专家调研模块	69
六、	用户调研模块	70
第三节	基于全媒体大数据的视频内容评估模型具体构成	72
一、	全媒体收视模块的构成	72
二、	全媒体社交舆情模块的构成	75
三、	全媒体传播力模块的构成	77
四、	专家调研模块的构成	78
五、	用户调研模块的构成	79
第四节	本章小结	79
第三章	内容评估体系模型实施——以内容银行内容评估 体系为例	81
第一节	数据库设计：基于 MongoDB 进行架构	81
一、	数据库的选型：MongoDB	81

二、数据库的具体构成	83
三、建立传媒领域专业词典作为后续分析的基础	83
第二节 数据采集和预处理	86
一、通过爬虫和 API 采集开放数据	86
二、问卷系统采集分析师及用户调研数据：灵活、按需分配的 问卷系统	93
三、预处理：过滤及从非结构化到结构化数据的抽取	94
第三节 文本信息挖掘	99
一、关键词提取技术	99
二、文本情感倾向性分析	103
三、文本话题聚类	110
第四节 指数计算	118
一、互联网上的主流排名算法	118
二、内容银行内容评估体系：借鉴多种算法，综合文本评估结果 进行 3 种量化计算	128
第五节 本章小结	145
结语	146
参考书目	148

绪论

第一节 研究缘起与背景

本书的核心主题是对基于大数据的全媒体内容评估理论及其模型建构的探索。之所以建立这样的研究核心主题,是因为长久以来笔者一直关注媒介产业,尤其是内容产业及各个产业链环节的发展。随着内容市场的发展,生产的无限化及需求的个性化不能顺畅对接,急需一个与内容产业相适配的交易系统。

黄升民教授于2008年提出内容银行概念,并与业内的相关机构进行合作,逐渐形成了丰富、完善的理论体系,并以科研项目的形式将这一理论体系进行了落地,建设完成了内容银行交易平台及评估系统。笔者亦参与其中的指数建设、开发等工作,在这一过程中逐渐发现了当前业界在内容评估上的缺失,并对大数据的具体运用、大数据与内容评估的结合有一定的思考与积累。

一、内容市场的变化

当前,内容产业不再是封闭的、单向度的,开放、共享共建、多边合作已经成为内容生产、传输、分发、交易中的普遍现象,此时,必须彻底改变内容产业的流程和运作模式,建立透明的交易系统。

针对于此,上海SMG、陕文投、北京东方雍和等都做了相关的尝试,包括渐次出现的版权银行、在线素材交易平台等产品。

在内容评估方面,伴随用户行为的多屏化、碎片化,传统的收视率调研已经不能适应制播方的需求而逐渐开始与新媒体融合,例如,索福瑞推出的微博收视率等;而基于数字载体的各类内容评估指数产品也越来越多,如视频网站推出的自有指数,酷云互动基于数字电视、互联网电视推出的大数据评估产品,尼尔森网联基于海量数据的收视率等。

二、学界的探索

学界对于如何评估内容价值一向非常关注,从20世纪八九十年代中国开始启动收视率调查开始,学界对收视率的研究就从未停止。

而在海外,影视产业发达,在影视内容评估方面走在前沿,如收视率、电影评估等都起源于美国、英国等。同时,海外传播学界对此的研究也在不断进行。

近几年来,伴随着互联网、移动互联网等新媒体的兴起,学界对于视频内容在多个渠道的传播链条、多屏之间的用户行为等均有相关的论述。此外,伴随着大数据这一概念的浮现,学界亦开始思考大数据对传统评估方式的影响和借鉴,以及在大数据背景下,如何更加全面、综合地评估内容价值。

三、长期的追踪性研究为本研究提供了支撑

中国传媒大学广告学院媒介研究所在媒体产业化、平台化的研究上具有较为深厚的研究基础,继2005年提出“信息平台”之后,在2008年提出了“内容银行”的概念,并在此基础上开始了理论体系与产业实践两个方面的研究。在2009—2011年,“内容银行”的科研团队,通过对国内外众多知名媒体机构、内容机构、营销机构的实地调研,创造性地借鉴银行的评估、交易、数据库建设和管理思路,完成了内容银行核心理论体系的建构。从2012年至今,该团队启动了内容银行系统开发工作,将理论研究推向系统实践阶段。这一系统中,我们选择内容评估作为切入点进行研发和理论建设。

与此同时,业界大视频的趋势愈加明显,视频内容行业的海量渠道、海量需求的现状也在对内容评估提出新的要求。在龙思薇博士的论文《内容银行内容价值评估体系的方法研究》中,初步提出了一套内容银行架构下的内容评估体系。

承接前人的研究基础,笔者深入参与了内容银行评估体系的实际技术开发,在历时两年的开发工作中,实际接触到了具体的数据采集、存储、挖掘、展现等全流程,与业界、学界的各方人士进行了较多的沟通,为本书的研究提供了强大支撑。

第二节 前人研究与文献综述

一、内容评估方面的研究

在中国,学界就内容评估体系的概念、构成、指标量化及计算方法、运作程序等问题进行了较多研究。

(一) 对于内容评估内涵方面的研究

“内容”,指事物所包含的实质性事物,其有三层含义:(1)物件里面所包容的东西;(2)事物内部所含的实质或意义;(3)哲学名词,指事物内在因素的总和,与“形式”相对。而本书的研究对象是视频内容,也就是指以视频形式表现出的各类内容。

“评估”,《现代汉语大辞典》给出的解释是“评价和估量”。针对不同的对象,有不同的评价、估量方式。

由于内容的含义广泛,所以对不同类型的内容评估方式各有不同,如版权评估、专利评估等都可列入内容评估中,因此,在前人的论文中,尚未出现对于内容评估一词的明确定义。但对于视频内容评估,则有相关的概念定义。

在中国,目前针对视频内容评估的研究主要是围绕着电视节目所进行的。之所以如此,是由于在视频内容领域电视是最早、最大的终端,在数字技术之前,电视台节目的制作、播出一直都是视频内容领域的主要研究对象。

陈英与刘自高认为“节目评估体系是对节目运作全过程的评估,包括播前评估、播后评估和播中监测”^①。这是从节目的播出方——电视台及视频网站的角度进行考量的。由于播出方需要投入大量资本进行内容的制作、采购,同时,传统上电视台、视频网站又直接面对观众,因此从播出方的立场进行节目评估的划分有一定的合理性。

在《电视节目评估体系解析——模式、动向与思考》中,刘燕南将节目评估分为节目质量评估和节目传播效果评估两部分,而节目传播效果评估又可以进一步分为预测性的播前评估和测量反馈的播后评估。这两种评估都会涉及对节目质量的评价,尤其在播前评估中,必然要对节目中的某些要素进

^① 刘桦. 基于“三维”视角的中国电视节目评估指标体系研究[硕士学位论文]. 湖南大学, 2010, 5.

行评估和测量；而播后评估则主要是将节目作为一个整体进行传播效果评价。^①在电视台的评估体系中，播后效果评估占大多数，对于播前的预测少有涉及。

郑兴东在收视率的基础上，探查了节目的相关要素（类型、播出的频道、时段等）对节目表现的影响；张君昌在《广播电视节目评估概论》中，对广播电视节目评估进行了系统的阐述；杨绮与张瑞华则对收视率、满意度进行了重点讨论。

在《中国内地广播电视节目评价指标体系研究——历史、现状与发展》中，柯惠新指出：节目本身涵盖了多种元素、多种变量，且具有丰富的层次，而节目的整个制播流程环节复杂，流程漫长，变量众多。因此，将节目作为对象进行评估，需要考虑这种特性，设定综合性的指标体系实现全方位的评估才能够准确反映、评判节目的客观情况。以此为标准审视国内的电视节目评价体系，会发现当前问题众多：首先，概念不清晰，包括对满意度、美誉度等的厘定；其次，多种不同指标如何进行统一也悬而未决。在实际的内容评估环节，相关机构的专业化问题、社会化问题，也需要在实践中给予评价。

（二）电视节目评估指标体系的研究

在具体的内容评估指标体系方面，业界的实践中收视率占据了评估指标的大半壁江山，而在学界，学者们针对这一现象进行了较多的分析探讨，他们一致认为，收视率不能全面衡量内容价值，因而应该建立更加客观的评估体系。

中国传媒大学柯惠新教授认为，考虑到内容产业不同环节的评估需求，节目评估应该由主体评估、客观评估和综合评估3个层次组成。具体的指标应该包括量化的收视率、满意度、定制化的专家评议和节目自身的成本4个方面。

在《电视节目评估：从量化分析走向质的研究》中，郑欣指出，单纯追求收视率是不全面、不客观的，电视节目应该将观众满意度纳入评估体系中，在评估中需要在量化分析之上加入质化研究。^②

刘燕南在《电视节目评估体系解析》一文中，界定、归纳了中国电视节目评估的概念并梳理了主要的评估模式，比较了当时各个电视台的评估体系，

① 刘燕南. 电视节目评估体系解析——模式、动向与思考. 现代传播, 中国传媒大学学报, 2011(1): 45~49.

② 郑欣. 电视节目评估：从量化分析走向质的研究. 南京师大学报(社会科学版), 2008(4): 45~51.

提出了由主观评价指标、客观评价指标、成本效益三部分组成的体系。^①

张海潮的《电视节目整合评估体系》是国内较早提出的基于电视节目生产运营全产业链流程的整合评估体系。该研究在总结国内外有关电视节目评估研究和应用成果的基础上,结合中国电视业的实际状况,构建出了一个新型的电视节目评估系统——电视节目整合评估体系。该评估体系根据电视节目生产运营中的设计—生产—播出—评估4个主要环节,设计出24个主、客观评估指标,通过一系列科学、严谨的评估方法和评估流程设置,使节目评估结果通过评估坐标图、综合评估指数、综合评估分析报告3个评估层次比较精确、清晰、完整地体现出来。^②

该评估体系既集国内外百分制打分法、市场评价法、综合评价指数法等现有节目评估方法之长,又有自己的发展创新。比如,令评估结果一目了然的评估坐标图的设计,4个节目运营主要环节中24个有关节目生产、播出和市场效果评估指标的全流程评估安排。

在构建电视节目整合评估体系的目标和指导原则的前提下,该作者设定了评估体系的整体结构框架和主要的评估环节。

在整体结构框架方面,该作者所构建的新型节目评估体系是由指针系统、评估系统、实施系统3个部分构成。其中,指针系统由评估指标、数据采集、计算方法3个部分组成,是电视节目整合评估系统的基础部分;评估系统由坐标图、综合指数、综合分析报告3个部分组成,是电视节目整合评估系统的核心部分;实施系统由相关设定和评估流程两个部分组成,是电视节目评估体系的执行部分。

在评估环节方面,该作者也构建了“电视节目设计评估”“电视节目制作评估”“电视节目播出评估”和“电视节目效果评估”4个部分。其中,电视节目设计评估主要是对节目是否选择了合适的题材、节目形态是否具有新颖性、其目标观众的状况如何进行考察;电视节目制作评估主要考虑的是在节目完成设计之后,制作者对主持人、演员的选择、电视表现手段、视觉和听觉效果应用等影响节目效果的电视表现手段的运用情况;电视节目播出评估主要评估检测的是播出平台对节目市场效果的影响;电视节目效果评估环节主要测

① 刘燕南. 电视节目评估体系解析——模式、动向与思考. 现代传播, 中国传媒大学学报, 2011(1): 45~49.

② 张海潮. 电视节目整合评估体系. 北京: 中国传媒大学出版社, 2009.

评考量的是节目的受众数量、质量,节目的投入、产出以及节目的发展潜力等情况。

在实施系统层面,该作者首先对评估小组成员的构成提出了意见。他认为,节目主创人员必须参与进节目评估小组,身份可以是制片人和栏目主编,也可以是主管领导,而且规定了其所占评估小组的人员比例。其次,还对被评估节目的相关资料、客观指标评估实施方法、主观指标评估实施方法以及其他考虑因素进行了阐述,并对评估流程的具体步骤进行了设定。

这几位学者在内容评估方面的研究是具有代表性的。我们可以看到,他们一致认为对于内容评估仅仅靠量化数据不能做到全面、客观和公平,因而应该考虑内容本身及制播流程的特殊性,设计综合性的指标和评估体系。

(三) 新媒体环境下内容评估的研究

随着新媒体的发展,内容从单一播出平台扩展到电视、PC、手机等多个播放终端,观众可通过多种渠道接触内容,因此,传统的收视率调研已经不能反映内容的真实情况了。在这样的背景下,学界对媒体融合环境下的电视节目评估也做了诸多探索。

中国传媒大学广告学院丁俊杰教授在《视网融合背景下的电视节目影响力评估体系创新初探》中认为,新的媒介形态是在三网融合推动下诞生的,同时,三网融合也促进了媒体经营业务层面的融合,包括营销模式、内容等。丁俊杰指出,这对传统电视节目评估体系提出了巨大挑战,尤其是收视率。评估的来源由电视端转变为多种新渠道和终端,而同时,评估主体不再仅仅是电视观众了,网民也必须受到重视。^①

样本量少,因而分析深度不足,难以形成对节目的全面判断。不能全面反映内容质量,这是传统收视率的重大缺陷,尤其在新媒体时代,这一缺陷尤为明显。北京大学新闻传播学院教授陆地在《电视节目评估体系的创建与创新》中指出了这一不足,同时提出了中国电视满意度博雅榜评估体系,希望建构一个内容全网数据评估系统。^②

还有部分学者试图在电视节目评估体系中纳入新媒体元素,对节目新媒体特征的某一方面进行评估。李岭涛、黄宝书在《网络影响力:中国电视的新

① 丁俊杰,张树庭,李未柠.视网融合背景下的电视节目影响力评估体系创新初探.现代传播,中国传媒大学学报,2010(11):99~102.

② 陆地.电视节目评估体系的创建与创新.南方电视学刊,2013(1):19~22.

型评价体系》一文中,主要从网络影响力的视角重新对电视媒体的影响力进行整体评价,其中主要包括3个指标:网络知名度、网络被关注度、网络收视度。^①

张树庭在《视网融合时代的电视节目评估》中,以视网融合背景下的电视媒体新生态为背景进行切入,对电视媒体与网络视频媒体的竞合关系、扩张路径进行了剖析,呈现出“竞合”融合是网络媒体与电视媒体两者关系的主流趋势。即为了应对来自网络视频的冲击,电视媒体应在继续固守大众市场、做强传统平台、提升自身竞争力外,积极推行多元化策略,取道网络新媒体平台,通过内容优势整合媒体资源,主动实现与网络视频的互相融合,在竞合中实现电视与网络化的生存与突破。与此同时,受众的网络收视行为日趋增多,网络讨论日益活跃,在个人博客、社区论坛中,网民对电视节目的讨论表现得十分活跃,各电视台的网络宣传工作也开始利用网络平台积极造势。因而,“唯收视率”的评价标准备受质疑,于是,该作者提出了电视节目网络人气指数体系(IPI)。“所谓电视节目网络人气评估体系,是对电视节目网络舆论状况的评估,以此考量电视节目在网络上的影响力。它既是电视节目自身影响力的一种体现,同时又具有延伸性,可通过网络舆论的传播对电视节目本身的收视率提升乃至品牌塑造产生影响。电视节目网络人气评估不仅表现为人气量级上的高低评估,同时,通过对网民讨论内容的分析,实现对网民态度和观点的甄别分析”^②。

具体到指标构成方面,电视节目网络人气指数体系(IPI)包括电视频道网络人气指数、电视栏目网络人气指数、电视剧网络人气指数、主持人网络人气指数以及舆论话题网络人气指数这5个单体指数。其中每个单体指数均由网络关注度和网络评价度构成,并涵盖新闻、论坛、博客、视频、Wiki等多种载体。网络关注度包括媒体关注度和网民关注度两个方面,并细分为参与度与波及度两个指标;网络评价度包括媒体评价度和网民评价度两个方面,其下又细分出多个源生指标和派生指标。

此外,该研究阐述了电视节目网络人气指数体系(IPI)的信源渠道选择和基于信源渠道的信源抓取、数据甄别及指数计算的操作流程。

① 李岭涛,黄宝书.网络影响力:中国电视的新型评价体系.现代传播:中国传媒大学学报,2008(3):127~130.

② 张树庭主编.视网融合时代的电视节目评估——中国电视网络人气指数体系理论、模型与应用.北京:中国广播电视出版社,2012,20~21.

总体来说,收视率这种单一数据已经无法满足内容在新媒体环境下的发展需求,不能为内容发展提供持续的动力,而新的、科学有效的评价体系亟须被提出,以促进内容产业的发展。

目前众多研究者已经指出了这种弊端和建立新评价体系的必要性,但缺乏对整个内容评估产品发展的历史性梳理。同时,他们所提出的新的研究模型主要集中于某一种新媒体特征,并不能全面采集多终端、多平台的数据进行建模。并且,这些研究大多数是定性的阐述,缺少实际数据的支持,相关数据分析较少,未能将新的评估模型与新的大数据的技术相结合,总体上略显薄弱。

此外,这些论述多停留在理论层面,并未进行实操,也因此很难预见到实际操作过程中可能遇到的问题,并给出具体的指导方向。

(四) 对本研究中的视频内容评估进行界定

从上文的整理可以看出,目前,对于视频内容评估有一定的研究和界定,但这种界定尚停留在以播出方,尤其以电视台为中心考量的阶段,未将更多的视频产业链条纳入其中,这与当前视频产业发展的现状有所脱节。

在内容一词的定义中,内容指“事物内在因素的总和”,对于任何内容而言,其内在因素确实影响其外在的形式和作为商品的价值。在视频内容产业中,诸如内容生产方、播出平台等亦会衡量视频内容各种内在因素的价值,进而判断内容本身的价值。结合前人研究中对节目评估的定义,笔者拟对本文的研究对象——“视频内容评估”作如下界定。

本研究中的视频内容是指以视频为表现形态的成熟的内容产品,其包括电视剧、综艺节目、电影、动画等。对于视频新闻、微电影、网友上传的短视频等视频内容,其在影视内容中体量较小,生产传播方式尚不成熟,不列入本研究的研究对象中。

由于视频内容的生产需要调动各类因素,而各类因素都对视频内容最终的价值产生影响,因此,视频内容评估必须要将各类影响其价值的因素考虑在内,包括导演、编剧、演员(主持人)等。若将视频内容视为一种商品,其生命周期包含了策划、投资、制作、采购、播出、播后等多个环节,每一环节牵扯不同的产业链条,不同链条又都有评估的需求,因此,视频内容评估又可以再划分为剧本评估、投资评估、播前预测、播中监测、播后评估等多个细分环节。

本书以此界定为基础进行研究,拟构建一种基于全媒体大数据的视频内容评估模型。

二、大数据的相关研究

由于本书是基于大数据的内容评估模型建构的,因此,梳理大数据这一概念的发展以及大数据的数据特征是文献整理的重要工作。

在传统的视频内容评估体系中,受制于技术以及数据处理手段的限制,一般都为播后评估,所以对于与“收视”相关的数据监测、指标选择等成为评估体系中的重点。但是,伴随着大数据技术与理念的出现、发展、成熟,在视频内容评估时可用的数据、数据获取的方式方法、数据处理的体系等方面都获得了全面的突破,所以传统视频内容评估体系不可避免地受到了一定的挑战。笔者在本书中亦是试图提出一种基于大数据的视频内容评估模型,因此,有必要对大数据进行界定。

(一) 对大数据概念和特点的研究

首先,数据分析是一直存在于各行各业当中的。经济规划、气象预测、物理学、基因工程等各个高科技门类都在应用大规模的计算分析。而伴随着信息技术的发展和应用,大数据这一概念被提出。

互联网、移动互联网、物联网都在飞速发展,人们几乎无时无刻不在贡献数据。据 IDC 估计,到 2020 年,人类社会将出现 500 亿个物联网传感器。在这样的背景下,数据的样态与此前发生了深刻的变革,无论规模、种类都大幅上涨,尤其是网页、图片及音视频等半结构化、非结构化数据在数据中占据越来越重的份额。数据量的增长态势也非常迅猛,据 MGI(麦肯锡研究院)预测,从 2009—2020 年,每年产生的速度量将增长 44 倍,达到 35ZB 的规模。而与此同时,数据存储、处理的成本和门槛都在下降。^①与 30 年前相比,数据存储的成本降低了百万倍,而诸如云计算、Hadoop、mapreduce 等技术的出现,使得大规模数据处理不再是高端行业的独有,而能够被各行业广泛应用。原本难于被处理的非结构化、半结构化数据,可以通过专门的数据库结构进行存储和处理。谷歌等企业在数据分析方面的应用,也使得大数据这一概念越来越进入业界、学界的视野中。

在 2008 年 9 月和 2011 年,《自然》和《科学》杂志分别推出了大数据的专刊,讲述了大数据在未来研究中可能的突破,并总结了数据在多个学科中愈发重要的作用。

^① 刘小刚. 国外大数据产业的发展及启示. 金融经济, 2013(9): 224~226.

麦肯锡的大数据相关报告《下一个创新、竞争和生产力的前沿领域》发布于2011年6月份,其在报告中阐述了数据挖掘的可能方法和潜在价值。

在2012年的达沃斯论坛上,《BigData, BigImpact》报告被发布。报告提出,数据将成为一种经济资产。

至今为止,关于大数据的具体界定尚没有出现广为业界和学界接受的版本,目前关于大数据概念的界定有以下几种。

(1) 麦肯锡:大数据是指其大小超出了典型数据库软件的采集、储存、管理和分析等能力的数据集。^①

(2) 维基百科:无法在一定时间内用常规软件工具对其内容进行抓取、管理和处理的大量而复杂的数据集合。^②

(3) Gartner:体量大、快速和多样化的信息资产,需用高效率和创新型的信息技术加以处理,以提高作出决策和优化流程的能力。^③

(4) Forrester:大数据本质在于“数据存储、处理和访问的流程与业务目标的集成”。

比较而言,后两者对大数据的定义更加具有普世性,不局限于数据本身,而是将整个数据流程囊括其中,提出数据的价值既来源于数据的“大量”,同时也得益于对数据处理分析方面的新的方法和工具。在这些概念的基础上, Teradata 从数据规模、数据种类等层面更加直观地描绘出了大数据是如何从传统数据演进而来的。

总体而言,上述概念都提到了大数据的几个重要特征:体量大、复杂、难以由传统的数据库软件所处理,对数据处理流程提出了新的要求。

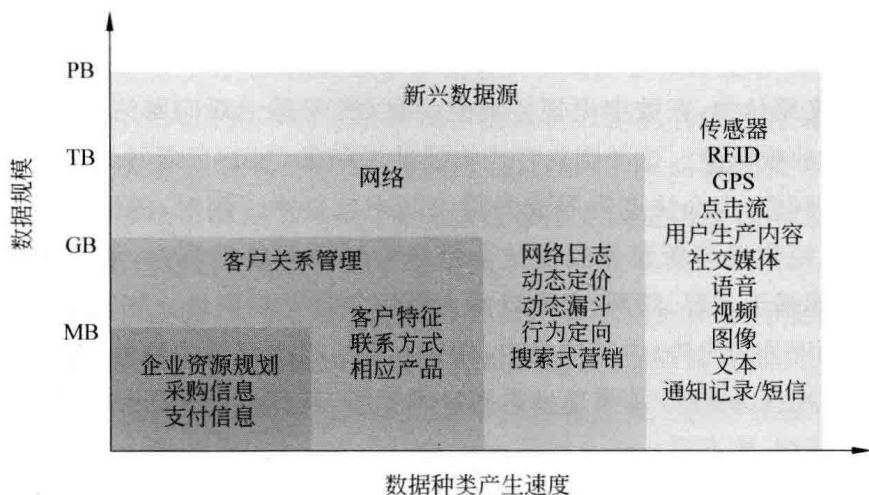
IBM于2010年提出了大数据的4V特征:海量的数据规模(Volume)、快速的数据流转和动态的数据体系(Velocity)、多样的数据类型(Variety)和巨大的数据价值(Value)。如笔者所述,内容评估体系的发展与数据密切相关,而数据的海量、动态、多样性和巨大价值均对视频内容评估体系提出了挑战,同时,也为评估体系的发展提供了现实基础。

由于数据量增大,不需要在少量样本数据这一镣铐的捆绑下去建构视频内容评估模型,而是可以拿到尽可能全面的数据。多样的数据类型和动态的

① 周云倩. 大数据时代的电视变局与因应之道. 中国电视, 2013(9): 90~93.

② 仇筠茜, 陈昌凤. 大数据思维下的新闻业创新——英美新闻业的数据化探索. 中国广播电视学刊, 2013(7): 12~14.

③ 袁冰. 大数据行业应用现状与发展趋势分析. <http://www.docin.com/p-1244164116.html>.



来源：Teradata、BCG

数据体系,意味着传统的视频内容评估数据流程要发生变化。大数据背后蕴含的巨大的数据价值,辅以合理的模型,将能够为视频内容产业带来推动力。

(二) 大数据在视频内容评估领域的应用研究

目前传媒领域对于大数据的应用日益增加,在这方面,也有多方研究。

在数据堂发布的《大数据产业调研及分析报告》中,对国内外的大数据产业发展链条和情况进行了整理,为本文在大数据产业方面的分析奠定了基础。

在《大数据时代对于传媒业意味着什么?》一文中,官建文认为,“大数据不只是一个概念,实际上是对一种社会状态的描述。谁拥有数据、掌握数据、主导数据并加以整合应用,谁就在社会中占据着重要地位”^①。

潘洪涛在《大数据背景下收视评估体系再思考》中提出,在新媒体和大数据环境中,能够采集的用户数据几乎是用户的总体数据,而非传统收视率调研的样本数据,这需要在统计学理论上作出一定的修正。总体数据的获得,可以克服统计学样本无偏、正态的前提假设,但是需要有更为高效的大数据处理方法。同时他还指出,虽然大数据为我们描绘出重构收视评估体系的美好蓝图,但是在目前情况下还有一些问题亟待解决,比如,家庭用户数据和个人用户数据整合问题、收视指标颗粒度大小问题等。

对于大数据应用于视频内容评估,有较多的研究集中在利用数字电视的

^① 官建文. 大数据时代对于传媒业意味着什么?. 新闻战线, 2013(2): 18~22