

·理论探索·

# 基于大数据的网络舆情分析方法研究

唐 涛

(上海社会科学院信息研究所, 上海 200235)

**〔摘 要〕** 随着互联网技术和应用的发展, 网络舆情在数据体量、复杂性和产生速度等方面发生巨大变化。网络舆情分析方法已超出了现有常用的分析框架, 必须在大数据分析的思维下有所创新。本文概述了常用的网络舆情分析方法, 归纳了当前网络舆情的大数据特征和分类, 提出了网络舆情分析创新要向大数据分析方向发展的观点, 并概述了 4 种新思路新方法, 对需要注意的相关问题进行了总结, 对未来网络舆情分析创新进行了展望。

**〔关键词〕** 网络舆情; 大数据; 舆情分析

DOI: 10.3969/j.issn.1008-0821.2014.03.001

**〔中图分类号〕** G203    **〔文献标识码〕** A    **〔文章编号〕** 1008-0821 (2014) 03-0003-04

## Research on Method of Internet Public Opinion Analysis Based on Big Data

Tang Tao

(Institute of Information, Shanghai Academy of Social Sciences, Shanghai 200235, China)

**〔Abstract〕** With the development of Internet technology and applications, network public opinion changed dramatically in the data volume, complexity and generate speed. Network public opinion analysis method has exceeded the existing analytical framework, and must make some innovative under the thinking of big data. The paper summarized common methods of public opinion analysis, induced the big data characteristics and classifications of public opinion, proposed the point that innovation of internet public opinion analysis should develop toward to big data analysis. The paper summarized four new methods and some issues which should be paid attention to, and forecasted the innovation trends of public opinion analysis.

**〔Key words〕** internet public opinion; big data; public opinion analysis

人类社会的数据产生方式主要有 3 种: 一是运营数据, 伴随着各类组织的运营而被动产生; 二是感知数据, 由遍布世界各个角落的传感器等设备监测社会运行和环境变化而自动产生; 三是人类原创数据, 由人单独创作和人际互动而主动产生<sup>[1]</sup>。其中, 人类原创数据在互联网, 尤其是 Web2.0 环境下产生了爆发式增长, 形成了庞大的网络数据资源, 而其中有相当一部分是由个人和各种社会群体, 对自己关心或与自身利益紧密相关的各种公共事务所持有的情绪、意愿、态度和意见<sup>[2]</sup>, 也就是网络舆情。

我国拥有世界上最多的网民和最大的访问量, 有效分析网络舆情, 对于政府, 可以更好地了解公众意见和诉求, 有助于建设安全和谐的网络环境和社会环境; 对于媒体,

可以突破传统“从记者到读者”的单向信息生产模式, 而加强对公众舆论的深层次分析, 实现新闻信息增值<sup>[3]</sup>; 对于企业, 可以掌握客户对产品与服务的评价及客户特征知识, 更好地提供个性化产品与服务, 实现利润增长。因此, 政府、媒体、企业和学术界都越来越重视网络舆情分析, 并不断追求分析方法的创新, 充分发挥舆情信息的价值。

### 1 网络舆情分析研究与实践现状

当前常用的网络舆情分析方法主要有网络调查方法、基于统计规则的模式识别方法、基于内容挖掘的主题监测方法等<sup>[4]</sup>。

网络调查方法是将社会调查法移植到网络上, 即在网

收稿日期: 2013-11-06  
基金项目: 本文系国家自然科学基金项目“移动互联网环境下网络舆情新特征、新问题与对策研究”(项目编号: 13CTQ028)研究成果之一。  
作者简介: 唐 涛 (1982-), 男, 助理研究员, 博士, 研究方向: 网络舆情、信息产业发展战略, 发表论文 10 余篇。

上进行问卷调查。这种方法通过设计问卷、抽样调查、统计分析等一整套科学程序，能够客观地推论社情民意。这种方法应用广泛，许多网站在新闻网页下方设置新闻评论功能和读者态度倾向调查，新华网、人民网等网站在近年全国两会期间专门进行关于民众关注热点的网络舆情问卷调查，一些网站还针对国家重大事件和社会热点进行网络调查。

在基于统计规则的模式识别方面，有学者通过统计分析某段时间内用户所关注信息点的记录，构建了互联网内容与舆情的热点/热度、重点/重度、焦点/焦度、敏点/敏度、频点/频度、拐点/拐度、难点/难度、疑点/疑度、粘点/粘度、散点/散度等 10 个分析模式和判据<sup>[5]</sup>。

在基于内容挖掘的主题监测方面，流程有 3 步：信息提取（包括信息采集、结构化数据存储）；信息预处理（包括信息过滤、词法分析、句法分析、概念分析）；舆情分析（包括文本标示、主题发现、意见挖掘、倾向分析），主要进行话题识别。近年来浅层分析技术出现，相关研究开始关注网络舆情的情感倾向。这种方法的核心技术包括搜索引擎技术、信息组织技术、自然语言处理技术等。

同时，市场上还出现了不少网络舆情监测分析软件，知名的有人民网舆情监测平台、拓尔思网络舆情监控系统、方正智思互联网舆情监控系统、邦富互联网舆情监控系统、军犬网络舆情监控系统等。以人民网舆情监测平台为例，网络舆情信息采集系统可对传统媒体网络版、新闻网站、论坛、博客等进行全天候定向抓取信息，还可利用百度、谷歌、奇虎等搜索引擎进行信息补充，并进行关键词、关注度、转载率等统计分析。但这些舆情监测系统擅长的是抓取新闻网页，在网络社区中，如 BBS、博客、微博、QQ 群、新闻跟帖等则效果有效，网络社区中的舆情主要依靠人工分析<sup>[6]</sup>。

在传统数据时代，我们分析舆情走势时，往往比较关注网民的言论，而忽视有多少人持此意见；往往重视解读文字内容，而忽视网民互动的社会关系网络。

## 2 网络舆情的大数据特征及分类

### 2.1 网络舆情的大数据特征

大数据是指无法在一定时间内用常规软件工具进行抓取、管理和处理的数据集合<sup>[7]</sup>，其在数据体量、复杂性和产生速度 3 个方面均大大超出了传统的数据形态，具有 4V 特征：规模性（Volume）、多样性（Variety）、变化快速性（Velocity）、价值（Value）。对当前网络舆情状况的观察可以看出，互联网的开放性使数量庞大的网民和各种社会群体可以在网上方便快捷地发表观点，这使得网络舆情的数

据量急速增长。其次，多媒体的发展使网络舆情的数据形态既有文本，又有图片、音频、视频等，呈现出多样性特征。再次，现代社会价值观念多元，各种观点交流交融交锋，舆论多元多样多变，网络舆情变化快速。各种因素共同作用，使得网络舆情数据越来越呈现出大数据特征。

### 2.2 显性舆情和隐性舆情

当公共事务发生后，网民在浏览、搜索、互动的过程中会随时发表言论信息，这些信息直接呈现了网民的情绪、态度、意见，可以被称为“显性”网络舆情大数据。

此外，还有一些数据并不是直接的网络舆情内容，但却从侧面客观反映了网民的关注热点、舆情主体之间关系等，可以被称为“隐性”网络舆情大数据。如，网民在浏览相关信息时，网站服务器端的日志记录了浏览页面 URL 等数据，在搜索相关信息时，搜索引擎服务器端的日志记录了搜索关键词等数据，可以反映用户的浏览偏好和搜索热点。网络社区用户之间互相“加关注”成为“粉丝”，服务器端记录的人际关系数据能够描述用户构成的社会化网络。用户之间互相转发和评论帖文所形成的转发量和评论量数据，可以反映帖文的重要程度。

## 3 网络舆情大数据分析方法的创新

网络调查方法主要是进行采样分析，其结果精确性随着采样随机性的提高而提高，与样本数量的增加关系不大，也就是说，样本的随机性比样本的数量更重要，但实现这样的随机性非常困难，以至于如果抽样对象是互联网用户这样的复杂和海量对象时，就很难找到一个“最优抽样”的标准，更不可能奢求抽样得到的小样本能够精确反映整体的所有特征。

基于统计规则的模式识别方法具有有效性，但由于不同信息源的信息产生方式有较大差异，该方法适用于对特点对象进行定点监测，有一定的局限性。

基于内容挖掘的主题监测方法主要针对“显性”网络舆情大数据，从现有的实践来看，由于受限于当前语义分析技术的精确度和速率，语义支持的缺失仍然是普遍存在的问题，一些工具难以有效地处理复杂的人类语言和不断变化的网络语言，而且对大规模数据分析的支持也较弱<sup>[8]</sup>，仍然需要大量采用抽样分析和人工分析。而人工分析又具有很强的定性化色彩，所以不少舆情分析报告经常使用“普遍表示赞成”、“不少网民认为”、“少数网民持反对态度”等定性化的语言，舆情分析的准确性难以进一步提高。

因此，要不断创新网络舆情大数据分析思路。一是绕开语义分析的技术瓶颈，开辟“隐性”舆情分析的“第二战线”；二是突破抽样分析的思维，用大数据方法分析收集

到的全体数据；三是将搜索数据、点击数据、人际关系数据、网民个人特征数据、相关社会数据等关联起来进行分析，深度挖掘出有价值的舆情。四是主动进行网络民意调查，有针对性地收集标准化数据。具体有以下几种方法。

### 3.1 基于网络日志数据挖掘的隐性舆情分析

当网民关注某公共事件而去浏览或搜索相关信息时，会在网站服务器端产生浏览日志或搜索日志。浏览日志中记录了网民 IP 地址、浏览时间、网页 URL 地址等数据，可以通过分析日志，统计某地区、某时间段内的浏览热点，许多网站推出的“舆情热点排行榜”就是这方面的应用。

搜索引擎后台的搜索日志记录了网民 IP 地址、搜索时间、搜索词、被点击的结果网页 URL 地址等数据。通过统计分析用户的搜索词及搜索频率，可以发现网民的关注点及其热度；对一段时间内与某个社会事件相关的搜索词进行词频统计，可以描述网民关注点的产生和变化过程。

目前，一些搜索引擎公司已经研发了通过搜索日志挖掘发现网络舆情的技术和应用。谷歌公司开发的“谷歌趋势”应用能统计某个关键词在一定时间段内某个地区被搜索的次数，将其与谷歌上随时间推移的搜索总量及当地的搜索总量相比较，得出该关键词的“相对搜索指数”，并将较长一段时间内的相对搜索指数描绘出来，以预测未来趋势<sup>[9]</sup>。一个成功的应用是谷歌制作发布的全球 20 多个国家的“流感趋势”。设计人员编入一系列与“流感”相关的关键词，包括“流感”、“温度计”、“发烧”、“咳嗽”等。当用户输入这些关键词，系统就认为可能与“流感”发病相关，继而跟踪分析并作出相对搜索量指数图。通过对以往指数的变化情况预测未来趋势，进而预报流感发病率。谷歌“流感趋势”在测试期间就表现出良好的预测效果，比美国疾病控制中心提前 7~10 天公布美国流感预报，且与官方公布的预报数据高度吻合，显示了基于搜索日志大数据进行预测的前瞻性和准确性。

### 3.2 基于社会网络分析的舆情主体关系发现

中国工程院李国杰院士认为，“数据背后是网络，网络背后是人，研究网络数据实际上是研究人组成的社会网络”<sup>[10]</sup>。互联网上不同主体间的互动形成很多社会化网络，以微博为例，用户之间互相关注、转帖、评论，假设用户乙关注了用户甲，则可以画一条由甲指向乙的有向边，表示甲发布的信息可以自动传递给乙。将所有用户之间互相关注的关系都画成有向边，整个微博舆论场就成为一个有向图，每个用户就是一个节点，每个“关注”就是一条有向边，所有人际关系数据最终全景展示了整个社会化网络。这些舆情主体间频繁联系、相互影响，在这个过程中涌现出一些威望和地位较高的舆论“意见领袖”，他们左右着其他主体的舆论方向，最终影响整个舆论场。同时，关注点

相似的舆情主体间也自觉或不自觉地形成了一些联系相对紧密的子群体，在子群体中信息传播速度更快。要管理和引导网络舆情，就必须对网络舆情主体和舆论子群体进行研究，而社会网络分析方法就是有效的手段。

“社会网络”的概念由英国人类学家布朗于 20 世纪 30 年代在研究社会结构时首次提出，到 20 世纪 70 年代，社会网络分析方法在社会学、心理学、人类学、数学、信息学等领域逐步发展起来。目前，社会网络分析方法已成为研究现实社会网络和以互联网为基础的网络信息交流的重要工具，其中就包括了个体中心度分析和凝聚子群分析。

个体中心度是评价一个人在网络中重要性的指标，主要包括点度中心度和中间中心度。其中，点度中心度用来衡量谁是网络中的重要人物。如果一个人可以将信息发送给更多其他人，那么他在网络中就拥有较大的话语权。因此，一个点的点度中心度可以用该点在表示网络的有向图中的“出度”来衡量。中间中心度衡量一个人作为媒介者的能力，即在网络中控制其他人的能力。如果一个人处于许多其他两点之间的路径上，则认为他具有控制其他两个行动者之间交往的能力。因此，一个人的中间中心度越高，就有越多人需要通过他才能与其他人发生联系<sup>[11]</sup>。凝聚子群分析主要揭示网络舆情形成者之间实际存在的或者潜在的关系，它们是否构成了相对较强的、直接的、紧密的或积极关系的小团体，这些小团体是否会成为促进舆情发展的核心群体。

根据舆情主体之间的“关注”数据，如果一个主体拥有的粉丝量越多，则他的信息能直接传递给其他人的可能性越大，他的点度中心度越高，他就越有可能成为意见领袖。如果一个主体链接的意见领袖的数量越多，则他越有可能成为传播信息的桥梁，他的中间中心度越高。还可以根据主体之间的相互关注数据，发现相互关注度高的子群体，他们之间信息的相互传播就越便利。

根据舆情主体之间的发帖、转帖、评论数据，如果一个舆情主体的原创帖文在一定时间内被转发和评论的数量越多，则他的点度中心度越高，那么他就是意见领袖。如果一个主体的转发帖文在一定时间内被再次转发和评论的数量越多，则他的中间中心度就越高，那么他就是传播信息的桥梁。同时，还可以根据主体之间相互转帖、评论的数据，发现互动紧密的子群体，他们之间舆情互动的实际效果更加明显。

当前，已经有一些成熟的社会网络分析软件，可以很好地进行社会网络分析，并呈现出可视化的分析结果，对于发现网络意见领袖和子群体有很好的作用。

### 3.3 关联不同领域数据进行舆情分析

大数据的一个重要特征是数据的混杂性，因此我们不

仅要接受多样化的数据，还要善于利用多样化的数据，将不同领域数据关联起来进行分析。

将用户职业数据、地域数据、年龄数据、专注领域等和社会网络数据结合起来，可以分析出不同的舆情热点在哪些职业、哪些地域、哪些年段、哪些团体中传播广泛，这对于更有针对性地进行舆论引导意义重大。

将网站新闻数据、论坛数据、博客数据、微博数据等进行对比，可以分析出舆情热点在不同舆论场的传播速度和广度，从而掌握哪些舆论场更易于传播哪类舆情。

还可以将舆情分析的数据与外部数据联系，如将食品安全问题舆情数据与相关食品的销售数据相联系，就能反映出舆情对企业经营的巨大影响。将网络谣言传播与造成的社会后果数据相联系，可以反映谣言的巨大破坏力；辟谣的引导性舆论发布后，再动态监测相关社会数据，可以看出舆论引导的效果。

### 3.4 基于网络民意调查的舆情分析

政府要进行舆情分析，只被动接受网络舆情数据是远远不够的，还需要走出去，主动收集数据，了解全社会对某项政策的评价。

现代意义上的民意调查实践起源于 19 世纪美国对总统大选的预测，发展到今天已经延伸到对各类社会现象的调查，且方法成熟。21 世纪后，我国才开始重视民意调查。2003 年，中国人民大学进行了第一个全国性的大型社会调查项目“中国综合社会调查”，将人们对社会热点事件和其他人群的看法作为调查内容。2006 年 9 月，我国才成立了第一个、也是目前惟一的国家级专门的民意调查机构——国家统计局社情民意调查中心。目前，我国互联网用户的人数、手机持有量都位居全球第一，这些都是网络民意调查的大数据来源。可见，我国缺乏的不是可供收集的数据和收集数据的手段，而是收集数据的意识。只有主动地进行网络民意调查，化被动分析为主动收集，才能更有针对性地满足对舆情的需求，也更准确地掌握舆情动向。

## 4 需要注意的问题

### 4.1 数据的安全性

在进行大数据分析时，隐私是个大问题。在大数据时代，数据的价值很大一部分体现在二次分析上，由于二次分析的结果未知，所以在收集数据前，进行“告知”和获得“许可”就不一定能起到绝对保险的作用。我们需要设立一个不一样的隐私保护模式，这个模式应该更注重数据使用者为其行为承担责任，而不是将重点放在收集数据之初取得数据所有者的许可。这样，数据使用者就需要基于其对数据所有者造成的影响，对涉及数据再利用的行为进

行评测。此外，我们还需要发明并推行新技术方式来促进隐私保护，一种途径是故意将数据模糊处理，促使对大数据的查询不能显示精确的结果，而只有相近的结果，这就使得挖出特定个人与特定数据点的联系变得难以实现并且耗费巨大<sup>[12]</sup>。

大数据的挖掘与利用应该有法可依。2012 年底，我国全国人大通过的加强网络信息保护的决策就是一个好的开始。现在很多机构和企业拥有大量客户信息。应当既鼓励面向群体、服务社会的数据挖掘，又防止侵犯个体隐私；既提倡数据共享，又防止数据被滥用。此外，还需要界定数据挖掘、利用的权限和范围。大数据系统本身的安全性也是值得特别关注的，要注意技术安全性和管理制度安全性并重，防止信息被损坏、篡改、泄露或被窃，保护公民和国家的信息安全。

### 4.2 数据的可获取性

要想提高大数据分析的效果，可获取的数据量是关键。大数据时代，传播形式的多元化使得数据分布在不同的传播终端，如一条新闻的网民评价可能分布在众多移动新闻客户端的用户评价中；再如微信中用户对社会事务的意见是一种处于半封闭圈群中的数据，非好友用户难以获取。这些问题都给设想基于大数据的网络舆情分析带来难题。除此之外，政府掌握着社会的大多数信息，而网络舆情分析还需要网站、运营商等企业的数据，所以政府和企业应该加强合作，共享数据。一些经过隐私保护处理后的数据还可以向全社会开放，以供企业、学术界和各社会群体进行研究，促进创新性应用。

### 4.3 数据的代表性

网络舆情大数据源于互联网的开放、共享，但我们不能仅仅凭借技术构建的大数据平台去打捞那些可能代表一个群体或一定数量级的“沉默的声音”。因为如果他们与网络隔绝，或者由于“沉默的螺旋”心理效应而不提供任何数据，那么我们所精心分析出的网络舆情实际上是不全面的，以此来指导社会管理是有偏差的。因此，我们在进行舆情分析的时候，网络舆情大数据只是一个方面，而非全部。

## 5 对未来的展望

随着网络技术和应用的不断创新，网络舆情在不断发生变化，面对数亿网民和浩如烟海的网络大数据环境，我们需要不断扩大网络舆情分析的内涵，从分析显性舆情内容，拓展到分析隐性舆情、舆情主体间关系、舆情子群体、舆情对社会的动态影响等更广义的舆情分析。我们要不断改进舆情的分析方法，网络舆情分析不再仅仅是采样分析，而是要开始关注大数据分析；不再仅仅是（下转第 11 页）

企业大数据能力，提高大数据工作水平。目前国内外有关厂商，纷纷推出与大数据服务相关的产品。如甲骨文公司推出了 Oracle 大数据机，最大程度帮助用户管理海量数据，快速提供一个具有高可用性的可扩展系统；IBM 推出最新数据库软件 DB2 10 和 InfoSphere Warehouse 10，能够轻松集成大数据系统，自动将数据压缩成更紧密的空间，以防止存储蔓延，并且将过去、现在和未来的信息进行分离，以清除某些程序代码。

#### 4.5 健全信息组织和信息制度

在大数据环境下，为了提高企业大数据能力，就必须要建立决策参谋式的信息组织结构，其业务流程要规范、科学，业务功能划分要明确，业务部门设置要合理，尤其要突出数据挖掘和智能分析在组织功能中的重要地位。同时，企业必须要针对大数据及其工作的特征，结合企业大数据工作实际，制定科学、合理、操作性强的系统化的企业信息制度，在具体企业大数据实践中，要严格按照相关制度来执行，加大制度的执行力。

#### 4.6 优化企业大数据支撑平台

企业大数据能力的提升是建立在企业先进大数据支撑平台基础之上，因此，现代企业要加大对大数据支撑平台的建设，构建先进的大数据技术支撑平台。优化企业的内联网和外联网，积极建设企业各种管理信息系统及集成化信息系统，积极开发和上线大数据技术产品，优化企业大数据支撑平台，为企业大数据工作提供良好的技术支撑，优化企业知识开发管理和利用能力。文献 [10] 指出：大数据核心分析能力的影响不仅存在于数据管理策略、数据可视化与分析能力等方面，从根本上也对数据中心 IT 基础设施架构甚至机房设计原则等提出了更高的要求。为了达

到快速高效的处理大量数据的能力，整个 IT 基础设施需要进行整体优化设计，应充分考量后台数据中心的高节能性、高稳定性、高安全性、高可扩展性、高度冗余，基础设施建设这 5 个方面，同时更需要解决大规模节点数的数据中心的部署、高速内部网络的构建、机房散热以及强大的数据备份等问题。

#### 参 考 文 献

[1] 李奕. 大数据应用方式：从数据服务、信息服务到知识服务 [N]. 中国计算机报, 2012—07—09, (24).

[2] 周渝, 唯奕. 大数据——企业运营中的新资本 [J]. 信息与电脑, 2012, (11): 19—21.

[3] 云计算环境下大数据及其智能处理技术 [EB]. <http://wenku.baidu.com/view/h20357b065ce050876321384.html>.

[4] 李佳师. 大数据：观念影响速度 [N]. 中国电子报, 2013—05—31, (5).

[5] 杨, 托尼. 利用大数据创造竞争优势 [N]. 中国计算机报, 2012—10—15 (23).

[6] 郭伟广. 改革企业营销模式，提高企业竞争力 [J]. 现代商业, 2012, (18): 27.

[7] 王海燕. 营销战略是企业竞争力的核心 [J]. 现代营销: 学苑版, 2010, (10): 16.

[8] 师大伟. 企业竞争力的核心——市场营销 [J]. 河北冶金, 2004, (1): 53—56.

[9] 甘绮翠. 大数据能力的关键影响因素 [J]. 销售与市场, 2013, (4): 22.

[10] 大数据的核心在于数据分析能力 [EB]. <http://www.stor-age.com/stor-age/2013/0111/2140090.shtml>.

(本文责任编辑：马 卓)

(上接第 6 页)

依靠语义分析，而是更求诸于自动化的数据分析；不再仅仅是对部分数据集的精确分析，而是涵盖多个数据集的关联分析。我们还要拓展网络舆情数据的收集方式，将分析系统自动产生的数据，发展到主动地、精准地收集舆情数据。为保障网络舆情大数据分析的安全和持续开展，我们还要解决可能涉及的隐私问题，以及推动数据的公开和共享。这些都将是面向大数据的网络舆情分析的潮流和趋势。

#### 参 考 文 献

[1] 孟小峰, 慈祥. 大数据管理：概念、技术与挑战 [J]. 计算机研究与发展, 2013, 50 (1): 146—169.

[2] 刘毅. 网络舆情研究概论 [M]. 天津：天津人民出版社, 2007: 51—53.

[3] 张嵘. 大数据时代舆情分析对媒体增值服务的可行性 [J]. 中国传媒科技, 2012, (11): 34—36.

[4] 陈忆金, 曹树金, 陈少驰, 等. 网络舆情信息监测研究进展 [J]. 图书情报知识, 2011, (6): 41—49.

[5] 谢海光, 陈中润. 互联网内容及舆情深度分析模式 [J]. 中国青年政治学院学报, 2006, (3): 95—100.

[6] 罗婷, 李成. 如何开发舆情监测产品——人民网舆情监测室的运作模式 [J]. 中国记者, 2010, (6): 52—54.

[7] Big data [EB/OL]. [http://en.wikipedia.org/wiki/Big\\_data](http://en.wikipedia.org/wiki/Big_data), 2013—04—26.

[8] 李广建, 杨林. 大数据视角下的情报研究与情报研究技术 [J]. 图书与情报, 2012, (6): 1—8.

[9] Google 趋势 [EB/OL]. <http://www.google.com/trends>, 2013—04—26.

[10] 甘晓. 大数据成为信息科技新关注点——访中国工程院院士李国杰 [N]. 中国科学报, 2012—06—27 (A1).

[11] 罗家德. 社会网分析讲义 [M]. 北京：社会科学文献出版社, 2010: 179—193.

[12] 维克托·迈尔—舍恩伯格, 肯尼斯·库克耶. 大数据时代 [M]. 盛杨燕, 周涛, 译. 杭州：浙江人民出版社, 2013: 220—223.

(本文责任编辑：孙国雷)