

大数据时代下服务业统计数据搜集方式的研究与优化

孙丽

辽宁省铁岭市清河区统计事务服务中心, 辽宁 铁岭 112003

DOI:10.61369/SDR.2025030016

摘 要 : 随着大数据时代的到来, 服务业统计数据的搜集迎来新契机和新挑战。传统统计数据搜集方法由于不精确、不及时等弊端已不能够很好地满足服务业迅猛发展的需要。文章阐述了大数据背景下服务业统计搜集数据受到的影响, 并分别对比了传统和大数据驱动下的服务业统计数据搜集方式之间的异同之处, 同时分析现阶段存在的问题, 并从各方面提出了优化方法来进一步提高服务业统计数据搜集的质量和效率, 以为服务业发展决策提供优质的数据支持。

关 键 词 : 大数据时代; 服务业; 统计数据搜集; 优化策略

Research and Optimization of Service Industry Statistical Data Collection Methods in the Era of Big data

Sun Li

Qinghe District Statistical Service Center, Tieling City, Liaoning Province, Tieling, Liaoning 112003

Abstract : With the advent of the big data era, the collection of service industry statistics faces new opportunities and challenges. Traditional methods, which are inaccurate and untimely, can no longer adequately meet the rapid growth of the service sector. This article discusses how the big data era impacts the collection of service industry statistics, compares traditional and big data-driven approaches, analyzes current issues, and proposes optimization strategies to enhance the quality and efficiency of service industry statistics, aiming to provide high-quality data support for development decisions.

Keywords : big data era; service industry; statistical data collection; optimizati

引言

数字经济时代服务业成为拉动经济增长的最重要力量, 在行业发展过程中其复杂多变性发展导致了统计数据搜集难度提高。传统的服务业统计数据搜集存在很多不足之处, 无法适应新的形势需要, 而大数据技术服务下的服务业统计数据搜集有所创新并且能够提供相关解决方案, 大数据服务背景下的服务业统计数据搜集对于服务业持续健康发展具有重要意义。

一、大数据时代对服务业统计数据搜集的影响

(一) 大数据技术概述

大数据是涵盖数据的采集、存储、处理、分析全过程的技术体系, 而物联网传感器、网络爬虫、API 接口等是实现数据自动采集和多源采集的技术方法, 它可以让智能设备采集到我们日常生活中使用的很多不同的方式的人群的各种各样的使用行为和行为记录; 还可以通过网络爬虫从网页中抓取公开的信息。数据存储是利用分布式文件系统、分布式数据库实现海量数据的高效管理和存储, 并且在集群环境下达到性能的最佳; 并且要保证支持对存放在集群环境上的 PB 级海量数据进行数据处理, 通常使用 MapReduce、Spark 等并行计算框架完成处理; 数据处理的数据分析使用了多种不同的方法, 比如有机器学习中的聚类分析、回归分析等, 以及用到了深度学习所使用的模型——神经网络模型、自然语言处理模型等, 都可以用于数据的数据分析中^[1]。

作者简介: 孙丽 (1976.1—), 女, 汉族, 辽宁铁岭人, 职称: 高级统计师, 学历: 大专, 研究方向: 统计。

(二) 大数据对服务业统计数据来源的拓展

大数据时代彻底改变了服务业统计数据的数据来源构成, 在大数据环境下主要是依靠企业报表数据、政府行政记录数据作为来源, 而大数据的应用使其数据来源数据呈现出井喷式的增长, 在互联网上有很多数据可以直接或者间接获取: 来自社交媒体上的大量用户自创的内容有消费者对于服务的态度以及服务的需求偏好和消费者的体验感受; 利用移动设备定位技术获取的位置数据可准确反映出消费者的消费轨迹, 并且反映出了服务业的空间分布与需求热点; 交易数据包含交易金额、时间、支付方式、交易频次等。比如: 共享出行行业产生的平台订单数据、平台行驶轨迹数据以及用户评价数据都分别展示了行业的全方位运行情况, 并且有资料可以做相应的证明。

(三) 大数据对服务业统计数据质量的提升

大数据的特性为服务业统计数据质量的提高提供保障, 大数据具有海量性的特点可以无限接近总体数据, 可以大大降低抽样

误差；而大数据的多样性可以从多方面对服务业的运行情况做出反映，能够有效防止出现单一数据源的信息偏差；大数据的高速度可以保持数据实时更新，快速抓取服务业的变化情况，有助于保证统计结果的时效性；大数据技术中数据挖掘算法能自动查找并剔除异常值和缺失值，利用数据检验、插值补全、逻辑校验等方式提高数据准确性，在数据分析上运用各种算法模型，能够更好地进行数据分析运算^[2]。——采用餐饮业交易数据、餐饮服务消费者评价数据、餐饮业人流量数据融合分析的方式预测餐饮业发展趋势，对于餐饮业发展运行情况进行有效跟踪和把控，更加科学客观地掌握餐饮业的发展趋势。

二、传统服务业统计数据搜集方式及其局限性

（一）传统统计数据搜集方法

通常传统服务业统计数据搜集方法包括：抽样调查、全面调查以及重点调查。抽样调查指用一定的科学方法从总体中选择一部分样本作为样本单位来搜集资料，并据以对总体特征作出估计，其优点是只用少量的调查工作就可以反映现象总体的基本情况；全面调查即是对总体的所有单位都进行调查了解的情况；重点调查是指调查总体中数量较多的单位。这种方法只是选择出了部分的重点单位，相对而言，抽样调查费时较少，同时又可以准确地得出需要的数据^[3]。通过对重点单位的调查，获取局部信息也可反映总体情况。这些方式曾经在传统的服务业统计数据统计过程中起到了至关重要的作用。

（二）传统方式在大数据时代的不足

随着大数据时代的到来，传统的数据统计收集方式逐渐暴露出问题：一是时效性差。传统调查从设计方案、数据采集、审核汇总直至发布，耗时往往需要几个月，甚至一年以上，不能反映出服务业瞬息万变的市场情况；二是精确性差。由于采取的抽样调查样本可能存在样本选择上的偏差问题（比如所选样本并不是为了服务行业里那些最相关的企业），其结果也无法被很好地利用来修正参数或者得到可靠的预测结论；三是不能很好采集关于服务业本身的各类数据（如服务创新、用户感受、服务过程等非结构化的服务数据）。

三、大数据时代服务业统计数据搜集存在的问题

（一）数据质量问题

虽然大数据能提高数据质量，但是数据质量问题依然存在，在准确性方面，大数据存在一定来源的数据失真问题。比如：用户在社交媒体上对所发布的评论容易受到自身情绪及虚真假消息的影响；被网络爬虫抓取到的信息也有可能出现错误或过期等问题。在完整性方面，各个平台提供的数据可能会产生数据缺失的问题，例如：电商的交易数据无法提供商品详情以及用户购买原因的相关描述信息，不利于对人们的消费行为做更深层次的挖掘分析^[4]。对于一致性来说，没有统一的行业标准，各平台关于某项服务有其自己的定义、分类与统计口径。以酒店房型为例，不同

的旅游平台分隔不同的房型标准，数据整合后会产生矛盾。如此就使得大数据在服务业统计的应用效果大打折扣。

（二）数据安全性与隐私保护问题

大数据时代，服务业统计存在的数据安全和隐私保护难题。服务业统计数据包含用户身份、消费偏好、企业经营等方面的信息，一旦泄露会对个人权利、企业利益产生重大危害。数据传输、数据存储中存在许多安全隐患，传统的加密方式难敌各种各样的网络攻击手段。其次，数据在中心化储存的过程中由于被聚集在一个点上，也容易被入侵利用。再次，在隐私保护方面，大数据的多元性增加了隐私泄露的风险，不同的来源之间的数据经过交叉匹配之后很容易泄露个体隐私。另外，数据的使用与分享并没有建立起完备的监督体制，在权利义务没有明确情况下，会导致数据被过度采集和非法应用。最后，在数据跨境流动时，国家之间互不相通的法律法规和监管要求势必会让数据安全保障越发困难重重。

（三）技术与人才问题

大数据技术在服务业统计数据搜集方面面临技术和人才短板两大制约因素。一方面，在技术层面，部分地区的信息基础设施较为薄弱，尚不具备满足大数据存储、处理、分析等方面的要求；有的市县级统计部门存在服务器性能低、缺少专业的大数据处理软件等问题，无法完成海量数据的处理任务；另一方面，在人才层面，既懂统计学理论知识又熟悉大数据技术的相关人员较为匮乏，传统的统计人员对大数据技术不了解，不能够将大数据技术运用于统计工作中；大数据技术人才不了解服务业统计的业务情况，不能从实际业务上把大数据技术运用到服务业统计中^[5]。因此，大数据的这些人才结构问题会制约大数据在服务业统计中的应用。

（四）数据整合与共享问题

服务业统计数据来源多样，不同的数据来源有不同的数据格式、编码规则、存储方式，如果要将来自社交媒体的文本数据、企业数据库中的结构化数据、传感器采集到的实时数据等实现汇聚，则需要经历较为复杂的检验过程、转化过程及标准化的过程。再者，由于数据共享机制不健全造成数据孤岛现象严重，跨部门、跨行业政府之间，政府与企业之间，企业与企业之间存在数据壁垒，没有建立统一的数据共享平台以及相应的数据共享激励机制，致使数据不能在不同的主体之间自由流通和共享，不能发挥大数据共享的价值。

四、大数据时代服务业统计数据搜集方式的优化策略

（一）建立大数据与传统方式相结合的搜集体系

建立大数据与传统统计方式相辅相成，协同发展的数据采集体系是优化服务业统计数据采集的主要手段，在传统统计指标数据采集上坚持采取规范稳定的传统方式，保障基本数据真实可靠，在采集反映服务业新业态、新模式及实时动态变化的数据过程中，充分利用大数据采集的优势；在统计零售业营业额时，企业报表可以得到全部总体销售数据，利用电商交易大数据统计分

析不同品类、不同时段的销售变化及消费者偏好；在开展抽样调查中，运用大数据构建更好的抽样框，利用用户行为数据、企业经营数据等多元化的数据作为辅助，扩大样本量并提高样本代表程度；两者相结合，互补长短，使得数据资料更加全面、及时。

（二）加强数据质量控制

构建全过程的大数据质量管控体系，保障服务业统计大数据的应用，是大数据服务服务业统计的基础工作之一，在数据采集过程中要建立健全数据源评估机制，对数据源从来源是否权威、是否可靠、是否完整等方面进行评估，优先选择好的数据源^[6]。与行业权威机构、大型企业合作获得高质量数据。对于大数据，使用数据挖掘算法及机器学习模型进行清洗、预处理，并发现错误数据，更正错误数据、补齐缺失值、统一只一数据格式，建立大数据质量监测指标体系，对数据准确性、完整性、一致性等情况进行实时监测和评价，一旦出现数据质量问题，利用大数据质量追溯系统，找到问题源并及时整改，定期开展数据质量检查、测评等工作，对于不达标的及时进行回炉再加工，保证统计数据质量。

（三）完善数据安全与隐私保护机制

健全部门健全的数据安全和隐私保护机制，对于服务业统计的大数据应用具有重要意义。从法律层面来说，完善数据安全和隐私保护的立法工作，明确规定数据收集、存储、利用、分享等的权责，规范数据处理行为；从技术层面来说，运用加密技术、访问控制技术、数据脱敏技术等全方位保护敏感数据的全生命周期，并对数据传输过程中的数据进行加密传输，对数据进行存储时使用加密存储，对一些涉及个人信息及企业商业机密等重要的数据做脱敏处理，此外还可以根据实际需要构建数据安全应急响应机制，制定数据泄露应急预案并定期组织预案的演练来应对突发的数据安全事件。加强对使用数据人员进行数据安全和隐私保护方面的培训教育，使使用数据人员增强安全意识，严格按照相关法规规定要求操作处理数据^[1]。

（四）提升技术水平与培养专业人才

提升技术、加强专业人才的培养是大数据运用于服务业统计的核心要素。在技术层面要加强信息技术基础设施建设，搭建高

效的大数据存储、处理平台，采购高效的大数据处理软件及工具；可以鼓励科研机构、企业加大大数据相关技术研发力度，突破数据采集、校验、分析等关键大数据技术，提高大数据处理的效率与准确性^[8]。在人才培养层面，高校、职校要对学校的专业课程设置做出相应调整，增设大数据与统计学融合的相关专业性课程。企业和统计部门可以从企业内部人手，为企业的在职职工开展相关的大数据技术培训及学习交流会，提高人员的技术水平，对企业内部人员进行有效的技术培训，并且建立完善的人才激励机制，多方面的调动人的一些主观能动性。

（五）促进数据整合与共享

打破数据壁垒，破除数据孤岛，实现数据整合和共享，是充分发挥大数据在服务业统计中协同作用的有效途径。一是完善数据标准和规范，统一数据格式、编码、元数据和统计口径等数据规则要求，形成规范统一的数据格式标准；二是建设跨部门、跨领域数据共享平台，为政府、企业以及高校和研究机构提供数据共享的服务；三是政府用政策引导、资金扶持等方式鼓励企业参与数据共享，并给予相关优惠政策，例如对于企业参与数据共享的工作给予一定的税收优惠或财政补贴等；四是健全数据共享安全机制，在保障数据安全的同时有序推进数据的流动及共享；五是强化数据整合与共享技术研发，搭建智能化数据整合工具平台，提高数据整合的质量和水平。

五、结论

通过大数据技术以海量数据获取优势为基础、多元数据来源为抓手，丰富服务业统计数据采集渠道，解决服务业统计数据时效性不足、全面性欠缺的问题；通过融合大数据与传统方式数据采集、加强数据质量管理、完善数据安全和隐私保护、建立平台化数据共享机制等方面举措突破大数据统计服务业过程中的难点问题，实现对服务业统计数据全方位的质量把控。提高服务业统计数据精准度、完整性和安全性，健全服务业统计体系，服务服务业相关政策的制订以及企业的发展战略决策。

参考文献

[1] 蔡媛媛. 试探讨“大数据时代” 规下服务业统计数据的搜集方式 [J]. 东方企业文化, 2015, (01): 189+193.
[2] 樊陈月. 大数据时代背景下统计数据分析方式的创新研究 [J]. 经济与社会发展研究, 2024(31): 7-9.
[3] 宋海峰. 试析大数据环境下的服务业统计构想 [J]. 数据, 2021, (08): 83-84.
[4] 付金叶. 基于大数据的服务业数据质量问题研究 [D]. 首都经济贸易大学, 2020.
[5] 黄小敏. 浅析大数据时代下统计数据的新认识与统计调查改进路径 [J]. 产业创新研究, 2020(20): 2.
[6] 李爱平. 大数据时代下的统计工作研究 [J]. 中国乡镇企业会计, 2020(4): 2.
[7] 魏佳. 大数据环境下服务业统计数据收集和整理 [J]. 现代经济信息, 2021(009): 147-148.
[8] 李鸿根. 大数据时代基层统计数据质量优化思考 [J]. 辽宁青年, 2023(24): 249-250.