

浅谈大数据时代如何做好数据统计工作

管小林

(汉中市南郑区统计局 陕西汉中 723100)

摘要:在社会经济发展的新时代,各类新技术和新型设备被广泛应用到各行各业中,对各行各业的生产效率、管理效率产生了较为积极的影响。其中,大数据技术的应用越发广泛,也逐渐被社会大众所熟知。大数据技术是一种综合性较强的技术,其基于现代化的信息技术和网络技术,是新时代社会产业升级的产物,并且其与现阶段人们比较熟悉的人工智能技术、AR 以及 VR 技术联系比较紧密,并且已经渗透到了消费购物等消费类型领域,对各行各业的生产加工以及日常维护管理和数据统计工作产生了较为明显的激励和推动作用。基于此,本文进一步分析了在数据统计工作中应用大数据技术的方式方法,希望可为广大从业技术人员和企业运营管理人员提供有效借鉴和参考。

关键词:大数据时代;大数据技术;数据统计;数据存储;数据预处理

引言

大数据技术的具体内容较多,数据工作的实现形式也比较多样,无论是数据采集、数据预处理、数据存储和处理,还是数据的分析和预测以及应用,均隶属大数据技术的应用范畴。从技术特点角度分析,大数据技术可一次性处理较大数量的数据,并且容错率高,可恢复性强。同时,大数据技术可对不同类型的数据进行处理,无论是文字文本、图片图像,还是音频视频,甚至是一些无规则、无实意的符号,均可成为大数据技术的“盘中餐”。此外,大数据技术还具有较好的应用密度性和适应价值,其应用安全性也比较高。技术人员在应用大数据技术开展数据统计工作时,应充分结合大数据技术的技术特点,并综合企业技术实现的具体要求,合理制定数据统计工作的内容和形式。

1、利用大数据技术提升数据采集的深度和广度

大数据技术为数据采集工作提供了较多的入口,但并非所有入口均是安全和有效的,技术人员在利用大数据技术对数据源进行采集时,应依据企业发展的实际要求和数据处理工作的密度要求,合理选择数据采集的入口,并科学应用数据发掘工具,提升数据采集的深度和广度。从数据采集的角度分析,物联网是数据采集的主要入口之一。具体而言,一套比较完整的物联网基础结构,可以比较清晰地记录用户使用物联网中各类基础设施的习惯,此类习惯将会以数据的形式存储在物联网中心服务器中,而此时,技术人员即可利用大数据技术发掘中心服务器中存储的用户各类服务类型数据,此类数据可为用户的睡眠时间、上班时间、用电量、用水量等。通过利用大数据技术,技术人员可提高数据采集的广度,针对单个体的多类用户行为进行数据统计,进而强化数据统计工作的体系性和完整性。更为关键的是,大数据技术为数据统计工作提供可选择的可能,即技术人员在应用大数据技术时,可这对单个体的单一行为数据记忆深度发掘,类如用户使用某一类 APP 的时长数据、用户操控某类家用电器的频率数据或者用户投资理财相关的数据等,从而在数据统计的深度层面,强化数据统计工作的现实效能。

2、利用大数据技术提升数据统计系统预处理能力

技术人员可利用大数据技术深入发掘广泛的用户数据,提升数据收集的广度和深度,但通过使用大数据技术收集的数据并非全都安全有效,数据的具体应用价值还需进一步统计和发掘,此时,就需要用到数据统计系统的预处理功能。从数据预处理的角度分析,技术人员在应用大数据技术时,应依据企业数据处理的具体需求,将数据的预处理功能分为数据类型统计功能、数据功用统计功能等,或者依据数据的用途将数据分为经济类分析数据、预警类分析数据或者科学技术类分析数据等。在应用大数据技术对数据进行预处理时,技术人员应控制好数据计算的内部规则,制定详实有效的系统预处理容错规则,提高数据预处理的有效性,并减小处理系统的运行压力。由于数据的收集过程一般均基于云端网络进行,数据的预处理过程也需要在数据传输和数据存储之前完成,导致数据处理中央服务器的短时运行压力可能过大,尤其在休息日、节假日或者其他特殊时期,数据量会出现激增的情况,进而影响数据预处理的效率,甚至出现“死机”的情况。为了避免此类问题的发生,技术人员可为大数据预处理系统指定分层式数据预处理规则,将数据按照处理难度分层,同时为数据处理加持一定的时间规则,从而引导数据处理系统可分等级、分时段对数据进行预处理,从而减少数据预处理的“拥堵”现象。

3、利用大数据技术做好数据存储和反馈工作

在数据被采集和预处理之后,就会被暂时存储到数据存储服

务器中。技术人员在应用大数据技术开展数据统计工作时,应结合企业数据类型、数据用途等特点,合理选择数据的存储形式,并在符合数据存储基本要求的基础上,尽量减少在存储数据过程中租用各类服务器的成本。具体而言,数据的存储形式可为单纯的数字,也可为音频、图片,或者视频。但音频、图片或者视频类数据的存储空间占用比例往往比较大,并且随着音频、视频清晰度、编码方式的不同,音频、视频的体积也会发生相应的变化。从此角度分析,技术人员在选择数据存储的形式时,应结合数据存储服务器的性能特点和数据存储的即时性要求,优化数据存储模式。同时,技术人员应合理使用大数据技术的实时反馈功能,基于网络技术,强化数据范围效果,从而为后续数据统计和分析工作提供有效参考,减少数据统计出错率,提升数据存储宽容度。

4、利用大数据技术强化数据定向输出工作

数据输出也是数据统计工作重要的一环,没有合理有效的输出方式,数据统计工作很难具有较好的可视化特性,此点对于企业经济分析或者社会问卷调查等统计工作尤为重要。技术人员在应用大数据技术开展数据统计工作时,应结合统计工作的特点和统计工作的实际目的,合理选择数据的最终显示形式,充分发挥大数据统计工作的可视化、可模拟化特性。具体而言,技术人员可利用大数据技术的数据检错和数据定位功能,筛选适合展示的数据。在此技术上,技术人员可利用大数据技术的可视化功能,将被筛选过程的数据以图表或者图形的形式展示出来,从而提升数据的可视化特性。从数据统计输出效果的交角度分析,技术人员可使用数据的可视化效果是否满足数据统计工作的预期来衡量数据统计工作的最终成效,并结合数据统计系统的反馈结构,分析数据统计模型的不合理之处,从而结合数据统计的现实要求,调整数据统计的具体方案,优化数据统计工作,此外,技术人员还应及时关注并分析市场环境的变化,将数据统计密度变化、数据统计价值量变化等因素加入到数据统计工作中,进而使大数据技术的应用过程更合理,也使数据定向输出工作更具体,更科学。

5、结束语

总之,本文针对在大数据时代如何做好数据统计工作提出了四方面的工作意见和建议,具体包括:第一,利用大数据技术提升数据采集的深度和广度;第二,利用大数据技术提升数据统计系统预处理能力;第三,利用大数据技术做好数据存储和反馈工作;第四,利用大数据技术强化数据定向输出工作。希望广大技术人员可结合企业经营的现实要求,合理选择大数据技术的应用办法,并灵活变通,进而可借助大数据技术的技术优势,使企业数据统计工作更具大数据时代的特点,从而提升数据统计工作的现实效能。

参考文献:

- [1]毛建成.浅谈大数据时代如何做好统计工作[J].统计科学与实践,2020(08):61-62.
- [2]齐金艳.浅析大数据时代如何做好卫生统计工作[J].商讯,2019(07):97-98.
- [3]齐金艳.大数据时代背景下如何做好卫生计生统计工作[J].大众投资指南,2018(15):247.
- [4]刘丽.大数据时代做好卫生计生统计工作刍议[J].科技风,2018(08):38-39.
- [5]杨永传.大数据时代做好卫生计生统计工作刍议[J].统计科学与实践,2017(09):61-62.
- [6]陈红梅.探讨大数据时代如何做好乡镇统计工作[J].中国国际财经(中英文),2017(11):98.