

电子信息系统中的数据采集与处理技术研究

孙晓辉 150404197310090059

摘要

随着信息技术的快速发展，电子信息系统已经成为各行业运行的核心支撑，数据采集与处理技术作为电子信息系统的基础环节，直接决定了系统的运行效率与信息输出质量。本文围绕电子信息系统中数据采集与处理技术的应用展开研究，首先梳理了数据采集与处理的基本概念与流程，分析了当前主流技术的应用特点，然后探讨了技术应用中面临的噪声干扰、数据异构性、实时性要求提升等问题，最后提出了优化数据采集与处理效能的方向，为电子信息系统的性能升级提供参考。

关键词：电子信息系统；数据采集；数据处理；信息技术

一、电子信息系统中数据采集与处理的内涵与价值

电子信息系统是指通过电子设备对信息进行采集、传输、存储、分析与输出的综合系统，广泛应用于工业生产、医疗健康、交通管理、金融服务等多个领域，其核心功能是实现信息的标准化流转与价值挖掘，而数据采集与处理则是整个系统运行的起点与核心基础。所谓数据采集，是指通过传感器、识别终端、网络接口等设备，从外部环境或其他信息系统中获取原始数据的过程，其核心目标是保证数据的完整性、准确性与及时性，为后续的信息加工提供可靠素材。

在当前数字化转型的背景下，电子信息系统承载的数据规模呈现爆发式增长，数据类型也从传统的结构化数据扩展到非结构化的文本、图像、音视频数据，数据采集与处理技术的价值愈发凸显。一方面，高质量的数据采集与处理能够提升电子信息系统的决策准确性，例如在工业智能制造领域，通过高精度的数据采集设备实时采集生产设备的运行参数，经过处理后及时发现设备的异常隐患，能够有效降低生产故障风险，提升生产效率；另一方面，优化的数据采集与处理技术能够降低电子信息系统的运行负载，通过在采集端实现初步的数据过滤，减少无效数据的传输与存储，能够降低系统的资源消耗，提升系统的运行稳定性。

二、电子信息系统中主流数据采集技术分析

当前电子信息系统中应用的数据采集技术，根据采集场景与数据来源的不同，

可以分为传感器采集技术、网络爬虫采集技术、接口交互采集技术与射频识别采集技术四类，不同技术有着不同的适用场景与特点。

第一，传感器采集技术是物理环境数据采集的主流技术，也是工业、环境监测等领域电子信息系统最常用的采集方式。该技术通过不同类型的传感器将物理量如温度、压力、湿度、位移等转换为可识别的电信号，再经过模数转换后转换为电子系统可处理的数字数据。近年来随着微机电技术的发展，传感器逐渐向小型化、低功耗、高精度方向发展。

第二，接口交互采集技术多用于不同信息系统之间的数据交互采集，是政务、金融等领域电子信息系统实现数据共享的主要方式。该技术通过约定统一的数据交互协议，如 RESTful API、SOAP 协议等，由采集方向数据提供方发起数据请求，提供方按照约定格式返回所需数据，这种采集方式能够保证数据的规范性与安全性，适用于结构化数据的定期或实时采集。

第三，网络爬虫采集技术主要用于公开互联网信息的采集，是舆情分析、市场调研等电子信息系统的主要数据来源。该技术通过模拟用户的网页访问行为，按照一定的抓取规则自动获取网页中的内容，并提取出所需的文本、图像等数据，经过格式转换后存储到本地系统中。当前主流的爬虫技术已经从静态网页爬取发展到动态网页爬取，能够应对 JavaScript 渲染的页面内容，同时也出现了多线程、分布式爬虫架构，大幅提升了大规模互联网数据的采集效率。

三、电子信息系统中主流数据处理技术分析

原始采集数据普遍存在噪声干扰、格式不统一、缺失值等问题，必须经过处理才能应用于后续分析，当前电子信息系统中主流的数据处理技术可以分为数据预处理、数据存储与数据分析三个环节，不同环节的技术特点各有不同。

在数据预处理环节，核心技术包括数据清洗、数据转换与数据集成。数据清洗主要是去除原始采集中的无效数据，包括填补缺失值、去除重复值、修正错误数据等，例如在传感器采集数据中，由于信号干扰容易出现异常峰值数据，通过基于统计学的异常值检测方法可以有效去除这类噪声，提升数据质量。数据转换则是将不同格式的异构数据转换为统一的格式，方便后续的存储与分析，例如将非结构化的文本数据转换为结构化的特征向量，将不同标准的编码数据转换为统一编码，解决多源数据的格式差异问题。

在数据存储环节，当前主流技术分为关系型数据库存储与非关系型数据库存储两类。关系型数据库如 MySQL、Oracle 等，通过二维表的形式存储结构化数据，具有一致性强、查询效率高的特点，适用于对数据准确性要求高的金融、政务电

子信息系统。

四、当前数据采集与处理技术存在的问题与优化方向

尽管当前数据采集与处理技术已经取得了长足的发展，但在电子信息系统的实际应用中，仍然面临一些亟待解决的问题。第一，噪声干扰与数据质量问题仍然突出，尤其是在复杂工业环境、复杂电磁环境中，传感器采集的数据容易受到电磁干扰、环境变化的影响，出现大量异常数据。第二，多源异构数据的整合难度较大，当前电子信息系统的数据来源越来越广泛，不同来源的数据格式、标准、精度存在较大差异，尤其是非结构化数据的特征提取难度较大，现有整合技术的效率仍然有待提升。第三，实时性与资源消耗的矛盾愈发突出，随着数据规模的增长，数据采集与处理需要消耗更多的计算与存储资源，而很多场景对数据处理的实时性要求越来越高，如何在有限的资源条件下提升处理效率，是当前需要解决的核心问题。

针对上述问题，未来电子信息系统中数据采集与处理技术的优化可以从三个方向推进。第一，推进边缘计算与数据采集端的融合，在数据采集节点完成初步的数据清洗与预处理，只将有效数据传输到核心系统，这样能够减少数据传输量，降低核心系统的资源消耗，同时提升数据处理的实时性。第二，优化异构数据处理算法，结合大模型技术提升非结构化数据的特征提取与整合效率，通过统一的语义框架实现不同来源数据的融合，降低多源数据整合的难度。第三，提升数据采集设备的抗干扰能力，结合自适应滤波算法在采集端实现噪声的初步过滤，从源头提升数据质量，降低后续数据处理的压力。

五、结语

数据采集与处理技术是电子信息系统的核心基础，其技术水平直接决定了电子信息系统的整体性能。随着数字化与物联网技术的发展，电子信息系统对数据采集与处理的质量、效率、实时性都提出了更高的要求，当前主流的采集与处理技术已经能够覆盖多数应用场景，但仍然面临数据质量、异构整合、资源消耗等方面的问题。未来通过边缘计算、人工智能等技术的融合应用，数据采集与处理技术将不断优化，为电子信息系统的功能升级提供更坚实的支撑，推动各行业数字化转型的深入发展。

参考文献

- [1] 王知非, 刘焕彬. 电子信息系统中数据采集及处理技术的应用[J]. 信息

技术与信息化, 2022(06): 138-141.

[2] 张宏科, 王鹏. 新一代电子信息系统数据处理技术发展现状与趋势[J]. 中国工程科学, 2021, 23(02): 145-152.

[3] 李刚. 工业电子信息系统中数据采集技术的优化研究[J]. 制造业自动化, 2020, 42(09): 156-158.