

DOI:10.19651/j.cnki.emt.2208887

基于多域并发的时间窗数据采集方法与应用*

郑惠泽¹ 丁克勤² 孙振山¹ 赵利强¹

(1.北京化工大学信息科学与技术学院 北京 100029; 2.中国特种设备检测研究院 北京 100029)

摘要: 针对工业数据采集中所涉参数繁多、数据结构复杂、采集频率高但数据质量难以保证等问题,提出了一种基于多域并发的时间窗数据采集方法,该方法基于滑动时间窗口的划分构建数据域模型,以多线程数据域并发的方法实时采集多源异构数据,针对高维度的多源异构数据进行扁平化处理,采用字节数据连接的方式对多源异构数据进行集成存储,实现了工业多源异构数据的采集与集成。通过仿真实验验证了该方法对于实时数据采集与集成存储的有效性,并在冶金起重机多维健康监测系统中应用该方法进行多源异构数据采集。实际运行结果显示,该方法能够提高冶金起重机多维健康数据的采集实时性,保证采集数据的质量,增加多维健康数据的集成度,是一种有效的实时数据采集集成方法。

关键词: 数据采集;多源异构;时间窗;数据域;数据集成

中图分类号: TP391 文献标识码: A 国家标准学科分类代码: 510.4033

Time window data acquisition method based on multi-domain concurrency and its application

Zheng Huize¹ Ding Keqin² Sun Zhenshan¹ Zhao Liqiang¹(1. College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China;
2. China Special Equipment Inspection and Research Institute, Beijing 100029, China)

Abstract: Aiming at the problems of many parameters, complex data structure, high frequency of acquisition but difficult to guarantee data quality in industrial data acquisition. A time window data acquisition method based on multi-domain concurrency is proposed. This method build a data domain model based on the division of sliding time windows. The multi-source heterogeneous data is acquired by the method of multi-threaded data domain concurrency and processed by flattening and connecting the byte data. Finally, the acquisition and integration of multi-source heterogeneous data is accomplished. The effectiveness of this method for real-time data acquisition and integrated storage is verified by simulation experiments. The method is applied in the multi-dimensional health monitoring system of metallurgical cranes for data acquisition. The actual operation results show that the method can improve the real-time performance of multi-dimensional health data acquisition of metallurgical cranes, ensure the quality of the acquired data and increase the increase integration of multi-dimensional health data. This method is an effective method of real-time data acquisition and integration.

Keywords: data acquisition; multi-source heterogeneity; time window; data domain; data integration

0 引言

工业设备的运行监测与健康诊断技术的发展,不断推动着数据采集技术的更新,同时大数据技术的发展又为工业大数据的采集提出了更高的要求。多设备分布式数据采集方式得到的工业数据时间不一致、结构分散、难以集成,很难满足工业大数据对多源异构数据集成的需求。对多源

信息采集处理有助于提高数据采集实时性,平化数据结构,降低数据结构的分散性,增强数据集成度,能够为大数据分析人员提供实时可靠的数据支持^[1-3]。

传统的多传感器数据采集方式采用分布式传感器、主机轮询式的采集方式,所采集的数据随着工业数据量的不断提升,由于采集设备的增多所造成的延迟会导致数据采集实时性与同步性受到影响。Song等^[4]将多节点分布式

收稿日期:2022-01-19

* 基金项目:国家重点研发计划课题(2017YFC0805103)项目资助

数据采集系统的命令延迟补偿与时钟恢复的相位补偿相结合,实现数据的高精度同步。Zheng 等^[5]提出了一种基于 FPGA 的数据采集与处理系统,数据可以在传输到 CPU 之前进行实时处理,在确保实时性的基础上能够以更高的采样率为数据采集提供数据。针对多源数据的采集从数据源层面^[6]、数据传输网络层面^[7]、采集上位机硬件与软件^[8]、软硬件处理过程^[9]等多个方面均做了不同的优化,以提高多源数据采集的质量。然而对于数据采集过程中数据解析造成的实时性影响并没有做进一步处理。

此外,工业现场环境恶劣,所安装不同类型的传感器受所在环境的高温、振动等因素的影响,不仅数据采集实时性较差,而且数据结构较为分散、不易集成。多源数据采集自多种不同的传感器,每种传感器所提供的数据均具有不同的性质,一般而言,多源数据中往往包含各种结构化、半结构化、非结构化类型的数据^[10-11],因此多种类型的数据往往会存储成不同格式的文件,从而导致数据分散,不易集成。现有工业数据集成方式主要使用关系型数据库与非关系型数据库结合的方式进行多源数据集成^[12-15],但是对于不同结构的数据仍是分散存储,不便于集中分析。

针对上述特点,提出了一种基于多域并发的时间窗数据采集方法,以多线程并发对多源数据进行采集与解析,以解决数据实时性的问题。同时对数据进行扁平化处理,以字节数据连接提高多源数据的集成度。

1 基于滑动时间窗的数据域模型构建

基于多域并发的时间窗数据采集方法依赖于在同一滑动时间窗内对多数据域的并发操作实现,因此要细化滑动时间窗口的划分与数据域的构建。

1.1 滑动时间窗口划分

滑动时间窗口用于对采集任务进行阶段性划分,时间窗口的大小是固定的,不固定的是所处的时间位置,其在时域上是可以滑动的。时间窗口的大小与各设备的采集频率紧密相关。一般而言,多源数据采集基于自身传感器特性与采集要求往往具有不同的采集周期,因此,要确定时间窗口的大小,必须要考虑全部的多源数据采集设备,以保证多源数据的完整性。

多源数据的采集周期是固定的,因此可根据其采集周期确定最小的时间窗口大小,设从第 i 个多源传感器获取到的数据采集周期为 T_i ,则最小的时间窗口大小可通过求取多源传感器采集周期的最小公倍数确定,时间窗窗口长度 T_w 为:

$$T_w = LCM(T_i), i = 1, 2, \dots, n \quad (1)$$

其中, $LCM(T_i)$ 为对多源传感器周期 T_i 求取其最小公倍数。

如图 1 所示为三台采集设备的采集周期与时间窗口划分。图中所示时间窗口利用 T_1, T_2, T_3 求取三者的最小公倍数进行确定,不同的设备在同一个时间窗口内具有不同

的数据采集量,最后将时间窗口内所采集得到的数据传输到数据域以备下一步的操作。

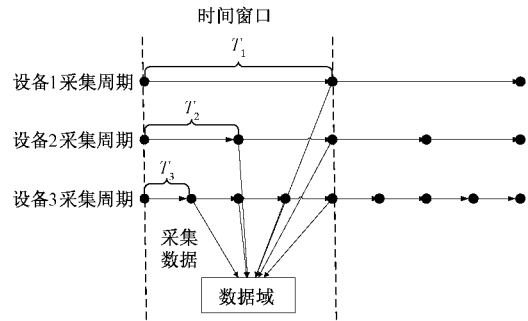


图 1 滑动时间窗口划分

1.2 数据域的构建

数据采集任务以时间窗口的长度将采集全周期进行阶段性划分后,在时域得到若干个连续的可滑动时间窗口。提取其中某段时间内的采集任务,以其所处的时间位置、数据类型作为域(Region)模型的时间与结构信息,并在其中添加该时间窗口中采集到的参数数据作为域模型的数据内容信息。因此域模型可作为数据采集时间段内某可测时间段以及其中所含数据信息的抽象表征。设 T 表征时间戳信息, S 表示数据结构信息, D 表示数据内容,则域 R 可以表示为 $R = \{T, S, D\}$ 。

在域模型的基础上进一步抽象化,以单独时间窗口的数据作为域的组成部分,得到数据域(Data Region)模型,数据域包含域的基本属性,同时可以被准确量化。设 R 表示域, T_w, S_i, D_i 分别表示每个时间窗口的时间信息、每个时间窗口内所采集数据的结构信息、每个时间窗口内采集到的数据内容,则数据域 R_D 可以表示为:

$$R_D = \{T_w, S_i, D_i\}, i = 1, 2, \dots, n \quad (2)$$

每个数据域均为可量化大小的数据包,其最主要的组成部分为数据头与数据体两个部分,数据头部分的数据主要包括数据域 ID、时间窗口信息、设备索引、结构索引等多种信息,具体意义如表 1 所示。数据体部分的数据则主要包括来自于各采集设备的原始数据及其数据结构信息表达。

表 1 数据头数据意义

信息名称	表达意义
数据域 ID	增量型顺序标志
时间窗口信息	所处时间窗口的起始与结束时间
设备索引	数据体中各设备数据所在处的索引
结构索引	数据体中各结构数据所在处的索引

如图 2 所示为单个数据域所包含的数据信息,其中数据体部分的数据按照设备分类与结构分类分别与数据头部分的设备索引与结构索引相对应以进行快速查询。

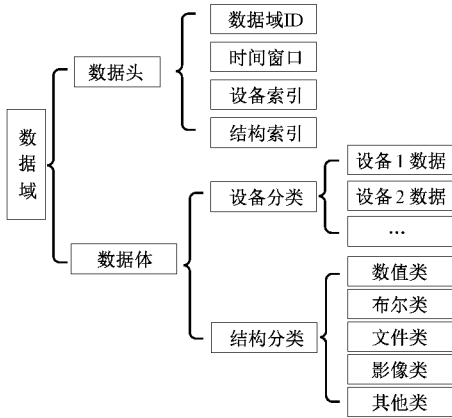


图 2 数据域结构

2 基于多域并发的时间窗数据采集方法

基于多域并发的时间窗数据方法采用多线程并行的方法,利用滑动时间窗口的思想,构造了多个数据域执行数据采集与集成存储。该数据域不同于传统意义上的数据处理区,不会在解析完成一次采集数据后才进行下一次采集,而是首先保证数据的采集,当阶段性采集任务完成后才转入数据解析,从而减少由于采集过程中解析延迟导致的数据采集缺失,这正是数据域所具有的优势所在。数据域是位于内存中的一个数据缓冲区,来自于多传感器的数据报文信息首先由数据域负责接收并做缓冲,同时利用多域并发的方法将不同的数据域进行功能划分,在同一时间窗口内分别进行数据的接收、解析与集成,完成多源数据的采集与集成存储。

2.1 多域并行的数据采集方法

多传感器采集得到的原始报文数据往往表示电流信号值或者电压信号值,在实际使用过程中,往往需要根据实际需要对所采集数据进一步进行量程匹配或数据运算方可得到所需数据。基于多域并发的时间窗数据采集方法将数据采集与数据解析分为不同线程,利用多个数据域并发的方式分别进行多设备数据的采集与解析,能够很好的减少由于解析造成的采集延迟。

以数据域作为基本数据处理容器进行数据的采集,首先开启主线程进行时间窗口的计算、数据域模型的构建等初始化操作,完毕后开启数据的采集与解析。数据域数据采集与解析采用多线程并行的方式进行,将连续时间窗口内的多个数据域利用不同线程分别完成各自的数据采集与解析,当线程被占用时开启新的线程与数据域,从而不断地在连续的时间窗口内进行数据域数据的轮转采集解析,进而将连续时间窗口内的数据域数据封装而成的帧数据连接成完整的数据流进行数据的存储。

如图 3 所示,时间窗长度确立以后,开启时间窗 T_{w1} 进行数据的采集,同时将采集到的原始数据放入数据域中,当该时间窗口到达截止期后,关闭时间窗口。在线程 1 中进

行数据域中数据的解析与发送的同时程序寻找空闲线程,即线程 2 将时间窗进行滑动并开启第二个时间窗,在线程 2 中进行下一时间窗数据的采集。如此利用多个线程控制不同数据域执行不同操作,进行并发处理以提高数据采集的实时性。

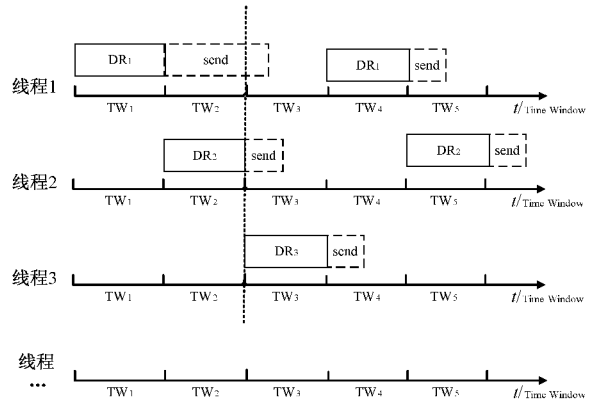


图 3 多域并行的数据采集方法

采集的数据将数据域信息作为连续的帧数据以数据流的形式发送到内存数据库中,同时内存数据库以数据域模型作为整体架构源源不断地处理数据流信息。形象地理解,内存数据库可认为是一个很大的数据域,每当内存数据库收到由数据接收端解析发送而来的一帧数据域数据时,内存数据库将其与之前的数据进行字节数据连接,并最终将得到的数据域数据进行输出,根据设定规则输出到二进制数据文件中以供后续调用。

2.2 基于数据域的多源数据集成存储方法

为数据域内解析完成后的数据设立数据分类规则,首先以参数数据的采集来源设备 ID 进行数据分类,其次依据数据结构形式对数据进行分类,包括数值类、文件类、影像类、布尔类及数字量的数据进行分类。同时根据参数采集设备与数据结构的不同创建不同的索引方式,创建数据域的操作方法,并根据数据域内时间窗口信息与数据结构信息生成数据域属性信息。其构建形式如图 4 所示。

图 4 所示的数据域构建形式,对包括数值类、布尔类、文件类、影像类等多种数据进行数据采集,同时对全部的采集数据以二进制形式进行数据的存储。每个数据域均包含数据头与数据体两个部分,而数据头则由数据体的索引信息与时间信息构成,因此在构建数据域的数据字节组成部分时,要先组成数据体,后组成数据头,最终再将其封装。

如图 4 所示数据体中每种设备所采集数据可分为 4 个部分,其具体意义如表 2 所示。

根据数据的类型,每种数据都有其单个数据最小的占用比特数,称其为数据元大小。根据数据元的大小即可计算每种设备在单个数据域中的实际采集数据量。

如图 4 所示,设数据元为 E ,实际数据为 D ,则该时间窗口内第 i 个设备实际采集的数据个数 N_{s_i} 为:

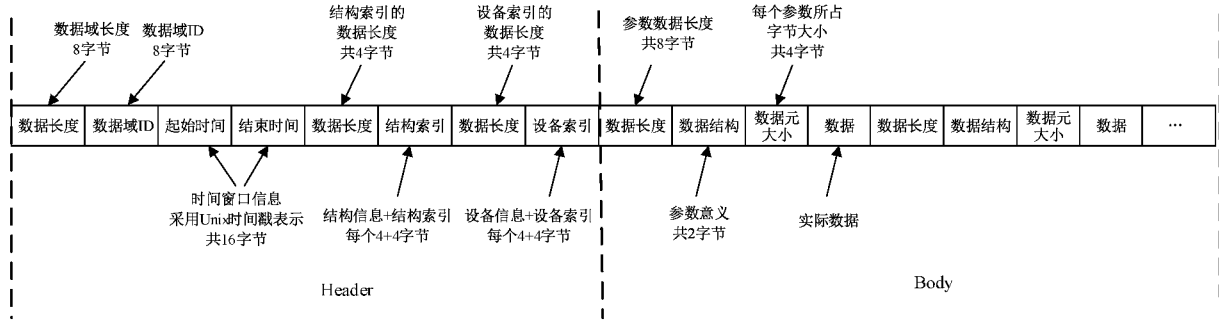


图 4 数据域构建形式

表 2 设备数据区意义

名称	意义
数据长度	数据域中该设备数据长度
数据结构	数据结构表示
数据元大小	单个数据所占最小比特大小
数据	实际存储的数据

$$N_{S_i} = \frac{Length(D)}{Vaule(E)}, i = 1, 2, \dots, n \quad (3)$$

其中, $Length(D)$ 代表实际数据的比特长度, $Vaule(E)$ 代表数据元的值大小。同时可根据第 i 个设备所具有的采集周期 T_i 与时间窗口大小 T_w 计算得到单个数据域中理论采集数据数量:

$$N_{L_i} = \frac{T_w}{T_i}, i = 1, 2, \dots, n \quad (4)$$

若 $N_{L_i} > N_{S_i}$, 则证明采集数据出现了数据缺失, 有部分数据在该时间窗口内未被完整采集。

传统的数值型数据分为 8 位、16 位、32 位、64 位整形、单精度、双精度浮点型等多种类型, 其主要的区别在于数据元大小不同, 因此在数据体每个设备的数据都专门使用一个 32 位整形常量来存储其数据元的值, 以便在解析时能够根据数据元大小进行数据的解析。布尔类数据每个数据只占用字节中的 1 位, 因此其数据占用最小字节数默认为 1, 不需要再在数据元中另行存储其大小。由于连续的 8 位布尔数据可以按顺序存储在同一字节中, 若不足 8 位布尔量, 则以 0 在后面补足, 因此在原本的数据元大小中标记其实际占用比特位数。

图像类数据可根据图像的存储方式以 RGB 三色图或灰度图进行存储, 并按照其二进制存储方式将图形按照二进制数据的方式输入到数据域中。如果是灰度图, 则根据图像的存储方式按照其分辨率以每个像素点一个字节 的值存储为二进制数。RGB 图形有红色、绿色、蓝色 3 种通道图像, 这些像素都具有从 0~255 的值, 其中每个数字代表像素的强度, 或者说红色, 绿色和蓝色的阴影。最后, 所有这些通道或所有这些矩阵都将叠加在一起, 这样, 当图像的形状加载到计算机中时, 它会是 $N \times M \times 3$ 字节的二进制数据, 并最终将其输出到数据域中。

如图 4 所示, 设第 i 个设备数据长度为 L , 数据结构为 S , 数据元为 E , 实际数据为 D , 时间戳为 $Timer$, 设备索引为 $IndexD$, 结构索引为 $IndexS$ 。

则单个数据域数据所占总字节数 D_{Byte} 可表示为:

$$D_{Byte} = 40 + 16n + \sum_{i=1}^n L_i \quad (5)$$

其中, L_i 表示第 i 个设备所采集数据占用字节数。

$$L_i = 14 + Length(D_i), i = 1, 2, \dots, n \quad (6)$$

通过对不同类型数据的转换与平化集成, 将其转换成二进制数据输出到数据域中进行数据下一步的解析与存储。

3 多域并发的时间窗数据采集方法设计与实现

对多域并发的时间窗数据采集方法利用具体程序设计了其程序执行流程, 并进行了程序的开发, 实现了多域并发的时间窗数据采集与存储的功能。

3.1 多域并发的时间窗数据采集程序流程设计

基于多域并发的时间窗数据采集方法利用多数据域同时并发的理念, 将数据集中不同层次的操作分散部署到不同的数据域中进行, 以减少多种操作造成的数据缺失及时间延迟。对于程序设计而言, 程序的执行要充分考虑到时间窗的关闭与多数据域的分配问题, 因此设计程序执行流程如图 5 所示。

如图 5 所示, 以多传感器为基础的采集设备接入采集系统后, 通过对多源传感器采集周期的匹配、计算, 得到相关的时间窗口长度, 对采集任务的全周期进行划分。同时建立数据域用于数据采集中的原始数据缓冲, 完成初始化操作。采集任务开始后, 位置最靠前的时间窗首先开启, 并利用 UDP、TCP/IP、Modbus、串口通讯协议、RSTP 等多种标准的工业数据通讯协议获取来自多源传感器的数据, 利用单个空闲的线程将得到原始数据报文源源不断的输入到建立好的数据域中。一旦到达时间窗口截止期, 则该线程数据域停止对原始数据的接收, 并开始对原始数据做进一步的量程匹配及转换等复杂性的操作, 直到完成后释放该线程。同时再寻找空闲线程或创建新的线程用于后续时间窗原始数据的接收。

基于上述程序执行流程, 使用状态机的编写理念对程

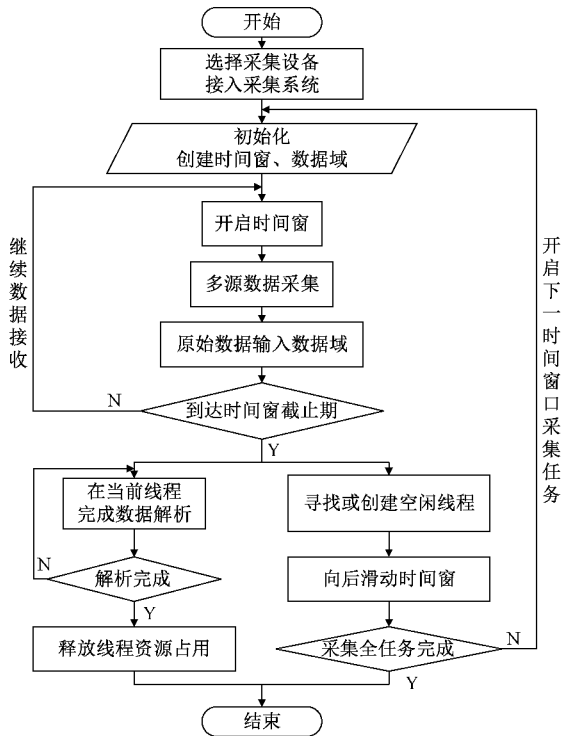


图 5 多域并行方法执行流程

序进行设计,利用多个线程的挂起、等待、忙碌、空闲、中止等不同状态的表示来进行线程的循环利用,从而将多个数据域数据进行并发处理,在不同的线程下实现数据的采集与解析,完成对工业多源数据的实时采集,状态机执行原理如图 6 所示。

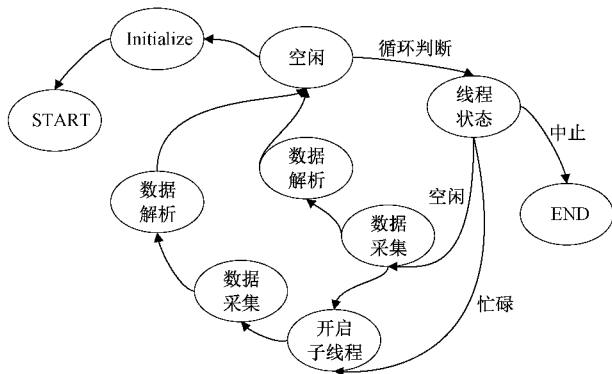


图 6 状态机执行原理

如图 6 所示为状态机基本执行原理图,状态机以事件作为驱动,执行相应的动作并完成状态间的切换。在基于多域并发的时间窗数据采集方法中,时间窗的开闭作为控制状态机状态切换的事件驱动,时间窗开闭会触发对当前全部线程的状态判断,并对线程的当前状态进一步判断,触发下一个状态,即利用当前空闲线程或创建新的线程,同时不同的线程中执行不同的数据操作。

3.2 多域并发的时间窗数据采集程序实现

LabVIEW 是一种图形化编程语言,具有很丰富的数

据采集、工业标准协议函数库,是开发数据采集、测量或控制系统的最佳选择。此外,在传统的基于文本的编程语言中,如果要创建多线程应用,必须编写代码创建多个线程并使之相互通信,而 LabVIEW 能够自动对 VI 进行判断,确定多线程技术应用的必要性,并由运行系统为用户处理线程间的通信^[16-18]。

基于多域并发的时间窗数据采集方法需要对多数据域进行操作,对于多线程技术的应用具有相当的必要性,同时需要使用不同的工业标准协议,LabVIEW 中的函数库能直接为其提供便捷的数据采集,因此使用 LabVIEW 进行程序开发能够加快开发进度,具有极大的便利性。

基于 LabVIEW 的 JKI 状态机与队列对多域并发的时间窗数据采集方法进行开发。采集程序分为数据采集层、数据输入层以及数据输出层。数据采集层利用多线程技术,使用工业标准协议对相应的数据进行实时采集,保证数据采集的实时性,同时将采集到的原始数据压入队列。数据输入层对多个解析线程进行标志位的判断,选择合适的空闲线程从队列中按照“先进先出”的原则将不同设备的原始数据取出,利用状态机的不同状态分别对各个设备的原始数据做相应的量程匹配或进一步解析等操作。数据输出层则对解析后的数据进行进一步的多源数据匹配与封装,进而对多源数据进行进一步的集成与存储。

程序首先利用两个单独的线程对原始多源数据进行采集,其次利用一个单独的线程利用线程状态数组控件的值实现对空闲子进程的选择与判断,在判断完成后根据条件结构将所需执行的空闲进程压入队列,令其执行数据采集操作。此外设置了 4 个内部操作完全相同的空闲子线程用作多域并发的数据采集与解析操作,并将其状态与线程状态数组控件相关联以实现主线程对子线程的控制。

工业多源数据包含的数据种类繁多,从传统结构化的数值类、布尔类的数据到非结构化的图像、声音、文件等多种类型的数据都有囊括。如表 3 所示,为方便数据存储与识别,对每种数据结构与意义做了相应的标注,通过对数据类型的判断即可进行相应的数据解析。

表 3 数据类型标识

数据标识	数据类型	数据元大小/bit
0x0001	Byte	8
0x0002	Short	16
0x0003	Int	32
0x0004	Long	64
0x0005	Float	32
0x0006	Double	64
0x0007	Boolean	1
0x0008	Char	16
0x0009	DTA 文件	...
0x000A	TXT 文件	...
0x000B	灰度图	...
0x000C	RGB 图像	...

如表 3 所示,对采集到的不同类型的数据按照其对应关系建立映射,根据设备增加或减少可适当调整表格。针对采集得到的数据,按照数据标识、数据元大小、原始数据的位数据进行连接,最后给出总数据长度插入到字节数据前面形成单个设备的数据域数据。此外,多源设备的全部数据采集完成后,按照多源设备的顺序将全部数据按照位数据进行连接,形成数据体,最后建立设备索引与参数结构索引,并附上时间窗信息,完成数据头的位数据连接,形成一帧完整的数据域数据。

利用 LabVIEW 中的循环结构对多源数据解析得到字节数据,首先利用数据源名称检索其数据元大小,并将得到的数据元值进行字节拆分组成为字节数组插入实际数据之前,其次计算解析数据长度,并与数据元所占字节、数据长度所占字节进行相加,最后将得到的值转换成字节数组与之前的数据进行数组拼接得到单个设备的字节数组,最终将多个设备的数据进行连接并构造数据头进行连接输出到二进制数据文件中进行存储。

4 实验与分析

4.1 仿真实验

为了验证基于多域并发的时间窗数据采集方法的性能,在 LabVIEW 中进行仿真实验。实验环境的硬件配置如下: Intel(R) Core(TM) i7-10750H CPU@2.60 GHz 处理器, 16 GB 内存, 64 位 Windows 操作系统。软件环境配置为: LabVIEW 环境, DSC 工具包, OPC Server、Modbus Slave 工具, 串口测试工具等。实验选择 OPC 通讯协议、RS232/485 串口通讯协议与 Modbus TCP 协议与 IMAQ 采集方式仿真不同数据的采集与生成, 实验测试参数如表 4 所示。

表 4 仿真实验测试参数

序号	采集方式	类型	采集周期/ms	数据元/Bytes
1	OPC 协议	Real	200	4
2	串口协议	Int	100	4
3	Modbus TCP 协议	Float	100	4
4	IMAQ	RGB 图	10 000	48 312

实验时采集时间窗根据多源参数的采集周期进行计算可以得到:

$$T_w = LCM(200, 100, 100, 10\ 000) \text{ms} = 10\ 000 \text{ms} \quad (7)$$

实验时将采集滑动时间窗的采集数量初定为 $N = 70$ 。实验采用 NI OPC Server、串口测试工具以及 Modbus Slave 工具、摄像头等工具对仿真实际数据的生成与采集, 仿真条件如表 5 所示。利用表 5 所示仿真工具对实际数据源进行仿真实验, 通过基于多域并发的时间窗数据采集方

法进行数据采集, 并通过设置“0xAABCCDDEEFF”的时间窗数据起始标识符将源源不断的时间窗数据以二进制形式存储到位于“D:\Multi_Source_Data”文件夹下的同一 DTA 文件中。本次实验主要是为了验证提出的基于多域并发的时间窗数据采集方法的采集性能, 因此分时间窗进行统计连续时间窗内采集到的数据量与理论值进行比较以验证数据采集的实时性。

表 5 仿真条件

仿真工具	采集量	仿真数据	仿真对应的实际数据
NI OPC Server	1	Real 数据	PLC 生成的 Real 数据
串口测试工具	3	Int 数据	AD 模块数据
Modbus Slave	5	Float 数据	应变采集仪采集数据
摄像头	1	132 * 122 像素图像	工业摄像头采集图片

首先对数据采集时间窗进行分析, 通过对比两次相同实验条件下采集数据时间窗的时间戳信息, 能够得到其采集时间窗受限于计算机时间精度会不断发生时间的漂移, 并在连续的时间内持续进行漂移, 其累积漂移时间如图 7 所示。

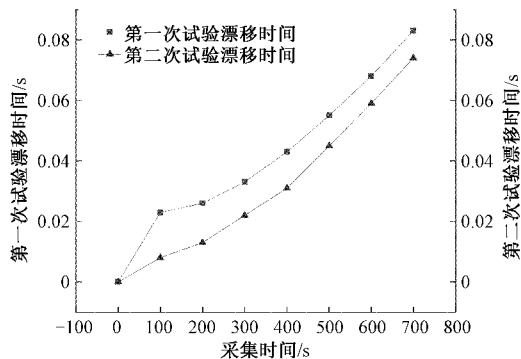


图 7 累积漂移时间图

如图 7 所示为数据采集过程中随时间增加所不断增加的累积漂移时间, 可以看出, 相同实验条件下两次实验其漂移时间增长的过程基本相同, 呈线程增长趋势。且在 700 s 的时间内仅漂移不超过 100 ms。对于采集周期为 100 ms 的数据采集过程而言, 若采集时刻正好在 100 ms 的开始或者末尾, 则有可能由于时间的漂移发生多采集一个值的情况, 因此采集数据量为 7 000 或 7 001 均为合理的。实验对比了 RS232/485 串口协议与 Modbus 协议两组不同采集方式的数据采集量与理论值, 最终实验结果如表 6 所示。根据实验结果能够看出, 由于时间漂移现象导致部分数据采集段的采集量与理论值不符, 但是通过 Modbus 协议所采集的总数据量为 7 000 个值, 与理论采集

量相符。通过 RS232/485 协议采集的总数据量为 7 001 个值,由于时间总漂移量为 0.074 s,接近于 100 ms,因此当采集时刻在 100 ms 的开始或者末端时,采集数据量受时间漂移的影响,会出现多采集一个值的情况,因此也与理论采集量相符。

如表 6 所示的数据来自于按照设计的数据存储方式对以时间窗分段采集并存储的多源数据二进制文件进行解析后得到的相同采集频率数据的统计量。从表 6 中可以看出,基于多域并发的数据采集方法在以时间窗进行划分的时间段数据采集上,能够达到与理论数据采集量相近或相同的数据采集结果,且总数据采集量与理论数据采集量相同,达到了数据采集实时性的要求。基于多域并发的时间窗数据采集方法能够在满足数据采集实时性要求的基础上完成了对多源异构数据的集成,证明了该方法的可行性与有效性。

表 6 实验结果数据统计

采集时间段	理论采集量	Modbus	RS232/485
0~100 s	1 000	1 001	1 001
100~200 s	1 000	999	1 000
200~300 s	1 000	1 000	1 000
300~400 s	1 000	1 000	1 000
400~500 s	1 000	1 000	999
500~600 s	1 000	1 000	1 000
600~700 s	1 000	1 000	1 001
合计	7 000	7 000	7 001

4.2 应用分析

在某冶金起重机多维健康监测系统上对该方法进行了实际应用^[19],如图 8 所示为某 2# 200/63 T 的冶金起重机,在该起重机上安装并测试了该多维健康数据采集方法。

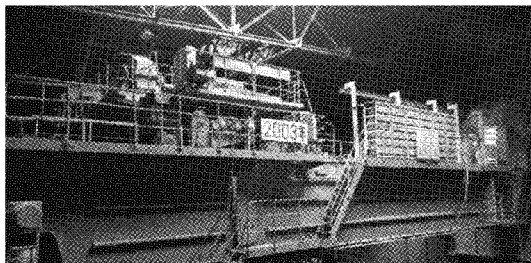


图 8 某 200t 冶金起重机

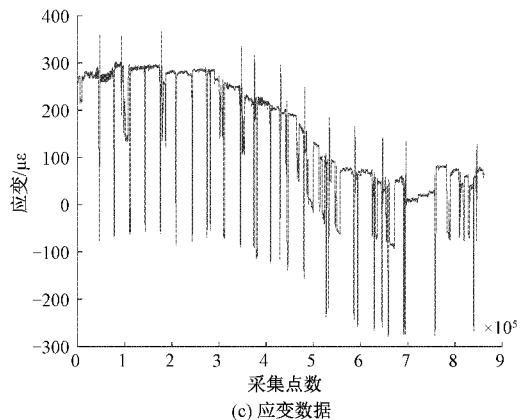
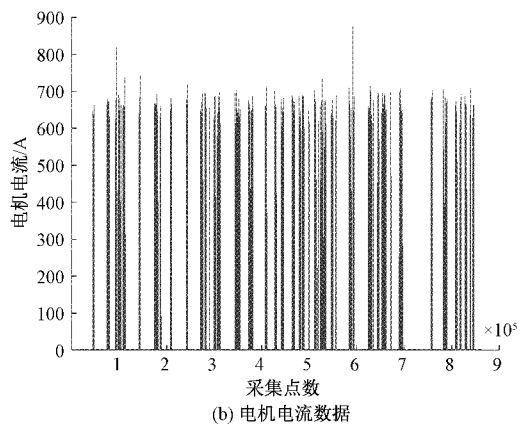
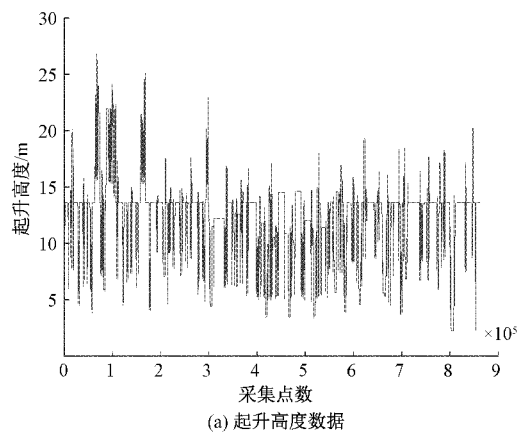
该冶金起重机最大起重量为 200 t,最大起升高度为 30 m,起升机构最大的速度为 12.3 m/min,车辆运行速度最大为 82.6 m/min。针对该冶金起重机运行与健康管理所必需的多源参数,利用多传感器对数据进行采集进行实验测试。

如表 7 所示为冶金起重机所需采集多源参数,在该冶金起重机上进行了软件的测试,软件在冶金起重机的实际运行过程中不间断运行进行数据的采集,同时将得到的数

据解析并按照日期时间的区别存储成二进制数据文件。在该冶金起重机的工作时间进行了实际的数据采集,采集所得数据如图 9 所示。

表 7 冶金起重机多源参数采集测试条件

关键采集参数	采集方式	数据类型	采集周期/ms
运行参数、电流参数、档位参数、限位参数	OPC 协议	Int Double Boolean	100
应力	Modbus	Double	100
振动	TCP/IP	Double	1
制动器位移	Modbus	Double	100
制动器行程图像	IMAQ	RGB 图像	10 000





(d) 制动器行程图像

图9 冶金起重机实际采集多源数据

如图9所示为加载冶金起重机实际运行存储的二进制数据文件所得,通过对实际数据的加载,证明了基于多域并发的时间窗数据采集方法对多源数据的集成的有效性。此外,冶金起重机运行在高温、高粉尘的环境中,在冶金起重机实际现场的运行结果显示,基于多域并发的时间窗数据采集方法对数据的采集具有极高的实时性,能够在极其恶劣的环境下保证数据采集的实时性与极高的数据质量,对于司机的操作与冶金起重机的状态变化能够做到及时有效的反馈。

5 结 论

本文提出了一种基于多域并发的时间窗数据采集方法,研究了该方法在工业多源数据采集中的有效性,在冶金起重机多维健康监测系统中进行应用并分析了该方法在实际冶金起重机环境下对数据采集、集成存储、显示的运行效果,具体结论如下:

基于多域并发的时间窗数据采集方法能实时准确的完成工业多源数据的采集,且能够保证数据的质量,解决了工业数据采集实时性较差与质量难以保证的问题。

基于多域并发的时间窗数据采集方法针对多源设备所具有多种结构的数据完成异构数据的平化集成,减小了数据的分散性,解决了工业多源异构数据集成度低的问题,为后续数据分析人员调用提供了极高的便捷性。

参考文献

- [1] 王春雨,蒋祖华,王福华,等.面向工业软件开发的半结构化知识语义检索方法[J].计算机集成制造系统,2021,27(8):2371-2381.
- [2] 李世昕,高红雨,苏航,等.基于关联算法的水泥熟料生产数据挖掘与分析[J].中国水泥,2020(2):73-79.
- [3] 王寅晨.分布式并行驱动环境下的参数化协同设计过程研究[J].机械设计与研究,2020,36(3):133-137,147.
- [4] SONG K Z, CAO G P, YANG J F, et al. A high-precision synchronous sampling approach for large-

scale distributed wire sensor networks in seismic data acquisition systems[J]. Instrumentation Science & Technology,2012,40(6):567-579.

- [5] ZHENG W, LIU R, ZHANG M, et al. Design of FPGA based high-speed data acquisition and real-time data processing system on J-TEXT tokamak[J]. Fusion Engineering and Design, 2014,89(5):698-701.
- [6] 于志强,陆桂军,于洪泽,等.组网式多通道数据采集系统设计[J].电气传动,2021,51(21):72-75.
- [7] JERECZEK G, MIOTTO GL, MALONE D, et al. A lossless network for data acquisition[J]. IEEE Transactions on Nuclear Science, 2017, 64(6):1238-1247.
- [8] JAVED M A, NAVEED M A, HUSSAIN A, et al. Integrated data acquisition, storage and retrieval for glass spherical tokamak (GLAST)[J]. Fusion Engineering and Design, 2020, 152:111424.
- [9] 张咏,吴琼,滕云田,等.激光干涉绝对重力仪数据采集与处理的时间优化方法[J].仪器仪表学报,2021,41(8):130-136.
- [10] 李仕峰.多源异构数据时空融合关键技术研究与应用[J].地理空间信息,2021,19(10):19-21,25,149.
- [11] 熊学锋,彭小庆,曹鑫.基于改进ORM的Oracle数据库异构资源整合方法研究[J].电子设计工程,2020,28(21):38-41,46.
- [12] AGHENTA L O, IQBAL M T. Development of an IOT based open source SCADA system for PV system monitoring[C]. Proceedings of 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), IEEE, 2019:1-4.
- [13] 刘晔,闫博,刘长鑫.基于厚板生产大数据的智能监控系统研究[J].控制工程,2020,27(12):2204-2209.
- [14] 李汉鹏,王阳,郑君旋,等.多源异构检测大数据采集与存储方法研究[J].电子质量,2021(9):53-55.
- [15] 黄欣,吴伟杰,张伊宁,等.电力系统多源异构数据混合式存储方式研究[J].机电工程技术,2021,50(1):77-79.
- [16] 武鹏飞,崔敏,王志斌,等.基于LabVIEW的暂态电能信号模拟及质量监测方法[J].国外电子测量技术,2021,40(8):41-45.
- [17] 徐天琪,杜磊,张殿龙.基于虚拟仪器技术的77GHz毫米波雷达目标运动参数模拟校准装置[J].电子测量与仪器学报,2021,35(8):118-125.
- [18] 秦程,任亮,王嘉健,等.基于LabVIEW的管道泄漏监测与定位系统[J].电子测量技术,2021,44(13):23-30.
- [19] 郑惠泽,丁克勤,孙振山,等.冶金起重机多维健康监测系统设计及实现[J].起重运输机械,2021(16):68-74.

作者简介

郑惠泽, 硕士, 主要研究方向为智能检测与传感技术、结构健康监测。

E-mail: Zhenghz_00@163.com

丁克勤, 博士, 研究员, 主要研究方向为先进传感技术、

结构健康监测、大数据分析、结构健康诊断等。

E-mail: kqding@sina.com

赵利强(通信作者), 工学博士, 副教授, 主要研究方向为智能检测与传感技术方向。

E-mail: zhaolq@mail.buct.edu.cn