



(21) 申请号 202311849146.9

G16Y 40/20 (2020.01)

(22) 申请日 2023.12.29

(56) 对比文件

(65) 同一申请的已公布的文献号

申请公布号 CN 117493777 A

CN 111770190 A, 2020.10.13

CN 117294531 A, 2023.12.26

CN 115514784 A, 2022.12.23

(43) 申请公布日 2024.02.02

CN 117236657 A, 2023.12.15

CN 103336510 A, 2013.10.02

(73) 专利权人 成都秦川物联网科技股份有限公司

CN 106933992 A, 2017.07.07

CN 110019488 A, 2019.07.16

地址 610000 四川省成都市龙泉驿区经开区南四路931号

CN 113361590 A, 2021.09.07

CN 113486063 A, 2021.10.08

(72) 发明人 邵泽华 李勇 吴岳飞 王峰

CN 113778967 A, 2021.12.10

CN 114203312 A, 2022.03.18

(74) 专利代理机构 广州三环专利商标代理有限公司 44202

CN 115080638 A, 2022.09.20

CN 115146712 A, 2022.10.04

专利代理师 周俊

CN 116384754 A, 2023.07.04

US 2023219228 A1, 2023.07.13

(51) Int. Cl.

G06F 18/10 (2023.01)

G06F 18/213 (2023.01)

G06F 18/214 (2023.01)

G06F 18/22 (2023.01)

G06F 18/241 (2023.01)

CN 112199363 A, 2021.01.08

审查员 孙玉梅

权利要求书3页 说明书9页 附图3页

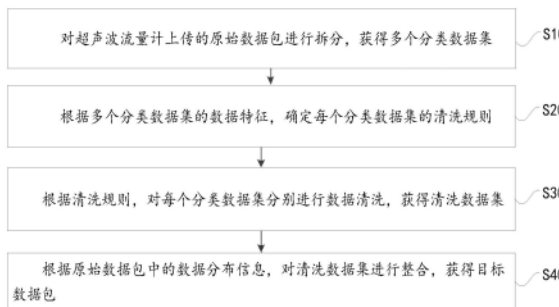
(54) 发明名称

基于物联网的超声波流量计数据清洗方法、系统及装置

(57) 摘要

本发明公开了基于物联网的超声波流量计数据清洗方法、系统及装置,涉及超声波流量计领域,通过对象平台将超声波流量计接入物联网系统,实现对采集的多源数据的处理,首先将超声波流量计上传的数据包进行拆分,得到多源数据的类别并将其分类得到分类数据集,然后根据分类数据集的数据特征,分别制定适合不同类别数据的清洗规则,分别进行数据清洗后,采集数据的质量得到提升,最后根据原始数据包中的数据分布信息,将清洗后的数据集整合为数据包,以便于再传输的质量,并且整合依据原始数据包中的数据分布信息,在目标数据包中还原真实数据的分布状态,能为后处理分析提供质量更好的

数据,有效提升后处理分析的水平。



1. 一种基于物联网的超声波流量计数据清洗方法,其特征在于,应用于物联网系统,所述物联网系统包括:依次交互的用户平台、服务平台、管理平台、传感网络平台以及对象平台,所述对象平台用于接入超声波流量计,所述基于物联网的超声波流量计数据清洗方法包括以下步骤:

对所述超声波流量计上传的原始数据包进行拆分,获得多个分类数据集;

根据多个所述分类数据集的数据特征,确定每个所述分类数据集的清洗规则;

根据所述清洗规则,对所述每个分类数据集分别进行数据清洗,获得清洗数据集;

根据所述原始数据包中的数据分布信息,对所述清洗数据集进行整合,获得目标数据包;所述根据所述原始数据包中的数据分布信息,对所述清洗数据集进行整合,获得目标数据包,包括:

根据所述原始数据包中的数据分布信息,获得所述清洗数据集的第一数据分布信息以及每个所述清洗数据集中数据的第二数据分布信息;

根据所述第一数据分布信息,对所述清洗数据集进行一次整合,获得整合数据;

根据所述第二数据分布信息,对所述整合数据进行二次整合,获得目标数据包;

所述根据所述第二数据分布信息,对所述整合数据进行二次整合,获得目标数据包,包括:

根据所述第二数据分布信息,获得每个所述清洗数据集中数据的时间序列排布;

根据所述时间序列排布对所述清洗数据集中数据进行二次整合,以使所述整合数据中各所述清洗数据集的序列长度的比值等于各所述分类数据集的序列长度的比值,获得目标数据包;

所述根据所述时间序列排布对所述清洗数据集中数据进行二次整合,以使所述整合数据中各所述清洗数据集的序列长度的比值等于各所述分类数据集的序列长度的比值,获得目标数据包之后,所述方法还包括:

判断所述目标数据包和所述原始数据包中数据的相似度是否小于相似度阈值;

响应于所述相似度小于所述相似度阈值,返回所述对所述超声波流量计上传的原始数据包进行拆分,获得分类数据集的步骤,循环至所述相似度不小于所述相似度阈值;

所述判断所述目标数据包和所述原始数据包中数据的相似度是否小于相似度阈值之前,所述方法还包括:

根据所述目标数据包和所述原始数据包,分别获得所述目标数据包和所述原始数据包中数据的分布相似度以及所述目标数据包和所述原始数据包中数据的序列相似度;

向所述分布相似度和所述序列相似度分配不同权重占比,获得所述目标数据包和所述原始数据包中数据的相似度。

2. 根据权利要求1所述的基于物联网的超声波流量计数据清洗方法,其特征在于,所述根据多个所述分类数据集的数据特征,确定每个所述分类数据集的清洗规则,包括:

根据多个所述分类数据集的数据特征,获得清洗数据类型、清洗手段以及清洗顺序;

根据所述清洗数据类型、所述清洗手段以及所述清洗顺序,确定每个所述分类数据集的清洗规则。

3. 根据权利要求1所述的基于物联网的超声波流量计数据清洗方法,其特征在于,所述对所述超声波流量计上传的原始数据包进行拆分,获得多个分类数据集,包括:

对所述超声波流量计上传的原始数据包进行拆解,获得多源数据;

对所述多源数据进行分类,获得多个分类数据集。

4.一种基于物联网的超声波流量计数据清洗系统,其特征在于,应用于物联网系统,所述物联网系统包括:依次交互的用户平台、服务平台、管理平台、传感网络平台以及对象平台,所述对象平台用于接入超声波流量计,所述管理平台包括:

拆分模块,所述拆分模块用于对所述超声波流量计上传的原始数据包进行拆分,获得多个分类数据集;

确定模块,所述确定模块用于根据多个所述分类数据集的数据特征,确定每个所述分类数据集的清洗规则;

清洗模块,所述清洗模块用于根据所述清洗规则,对每个所述分类数据集分别进行数据清洗,获得清洗数据集;

整合模块,所述整合模块用于根据所述原始数据包中的数据分布信息,对所述清洗数据集进行整合,获得目标数据包;所述根据所述原始数据包中的数据分布信息,对所述清洗数据集进行整合,获得目标数据包,包括:

根据所述原始数据包中的数据分布信息,获得所述清洗数据集的第一数据分布信息以及每个所述清洗数据集中数据的第二数据分布信息;

根据所述第一数据分布信息,对所述清洗数据集进行一次整合,获得整合数据;

根据所述第二数据分布信息,对所述整合数据进行二次整合,获得目标数据包;

所述根据所述第二数据分布信息,对所述整合数据进行二次整合,获得目标数据包,包括:

根据所述第二数据分布信息,获得每个所述清洗数据集中数据的时间序列排布;

根据所述时间序列排布对所述清洗数据集中数据进行二次整合,以使所述整合数据中各所述清洗数据集的序列长度的比值等于各所述分类数据集的序列长度的比值,获得目标数据包;

所述根据所述时间序列排布对所述清洗数据集中数据进行二次整合,以使所述整合数据中各所述清洗数据集的序列长度的比值等于各所述分类数据集的序列长度的比值,获得目标数据包之后,还包括:

判断所述目标数据包和所述原始数据包中数据的相似度是否小于相似度阈值;

响应于所述相似度小于所述相似度阈值,返回所述对所述超声波流量计上传的原始数据包进行拆分,获得分类数据集的步骤,循环至所述相似度不小于所述相似度阈值;

所述判断所述目标数据包和所述原始数据包中数据的相似度是否小于相似度阈值之前,还包括:

根据所述目标数据包和所述原始数据包,分别获得所述目标数据包和所述原始数据包中数据的分布相似度以及所述目标数据包和所述原始数据包中数据的序列相似度;

向所述分布相似度和所述序列相似度分配不同权重占比,获得所述目标数据包和所述原始数据包中数据的相似度。

5.一种计算机可读存储介质,储存有计算机程序,其特征在于,所述计算机程序被处理器加载执行时,实现如权利要求1-3中任一项所述的基于物联网的超声波流量计数据清洗方法。

6.一种基于物联网的超声波流量计数据清洗装置,其特征在于,包括处理器及存储器,其中,

所述存储器用于存储计算机程序;

所述处理器用于加载执行所述计算机程序,以使所述基于物联网的超声波流量计数据清洗装置执行如权利要求1-3中任一项所述的基于物联网的超声波流量计数据清洗方法。

基于物联网的超声波流量计数据清洗方法、系统及装置

技术领域

[0001] 本申请涉及超声波流量计领域,具体涉及一种基于物联网的超声波流量计数据清洗方法、系统及装置。

背景技术

[0002] 将超声波流量计与物联网连接,可实现对计量业务场景的全方位控制,通常为了便于传输,超声波流量计会将采集的流量、流速、气压等多源数据进行打包上传,物联网平台对数据包进行数据清洗后就没有其他处理手段,清洗后数据完全展开呈现,既不便于再传输,又打乱了上传数据中真实数据的分布状态,导致后处理分析的水平降低。

发明内容

[0003] 本申请的主要目的在于提供一种基于物联网的超声波流量计数据清洗方法、系统及装置,旨在解决现有技术中由于数据清洗手段的不完善,导致了后处理分析的水平降低的问题。

[0004] 为实现上述目的,本申请的实施例采用的技术方案如下:

[0005] 第一方面,本申请实施例提供一种基于物联网的超声波流量计数据清洗方法,应用于物联网系统,物联网系统包括:依次交互的用户平台、服务平台、管理平台、传感网络平台以及对象平台,对象平台用于接入超声波流量计,基于物联网的超声波流量计数据清洗方法包括以下步骤:

[0006] 对超声波流量计上传的原始数据包进行拆分,获得多个分类数据集;

[0007] 根据多个分类数据集的数据特征,确定每个分类数据集的清洗规则;

[0008] 根据清洗规则,对每个分类数据集分别进行数据清洗,获得清洗数据集;

[0009] 根据原始数据包中的数据分布信息,对清洗数据集进行整合,获得目标数据包。

[0010] 在第一方面的一种可能实现方式中,根据原始数据包中的数据分布信息,对清洗数据集进行整合,获得目标数据包,包括:

[0011] 根据原始数据包中的数据分布信息,获得清洗数据集的第一数据分布信息以及每个清洗数据集中数据的第二数据分布信息;

[0012] 根据第一数据分布信息,对清洗数据集进行一次整合,获得整合数据;

[0013] 根据第二数据分布信息,对整合数据进行二次整合,获得目标数据包。

[0014] 在第一方面的一种可能实现方式中,根据第二数据分布信息,对整合数据进行二次整合,获得目标数据包,包括:

[0015] 根据第二数据分布信息,获得每个清洗数据集中数据的时间序列排布;

[0016] 根据时间序列排布对清洗数据集中数据进行二次整合,以使整合数据中各清洗数据集的序列长度的比值等于各分类数据集的序列长度的比值,获得目标数据包。

[0017] 在第一方面的一种可能实现方式中,根据时间序列排布对清洗数据集中数据进行二次整合,以使整合数据中各清洗数据集的序列长度的比值等于各分类数据集的序列长度

的比值,获得目标数据包之后,方法还包括:

[0018] 判断目标数据包和原始数据包中数据的相似度是否小于相似度阈值;

[0019] 响应于相似度小于相似度阈值,返回对超声波流量计上传的原始数据包进行拆分,获得分类数据集的步骤,循环至相似度不小于相似度阈值。

[0020] 在第一方面的一种可能实现方式中,判断目标数据包和原始数据包中数据的相似度是否小于相似度阈值之前,方法还包括:

[0021] 根据目标数据包和原始数据包,分别获得目标数据包和原始数据包中数据的分布相似度以及目标数据包和原始数据包中数据的序列相似度;

[0022] 向分布相似度和序列相似度分配不同权重占比,获得目标数据包和原始数据包中数据的相似度。

[0023] 在第一方面的一种可能实现方式中,根据多个分类数据集的数据特征,确定每个分类数据集的清洗规则,包括:

[0024] 根据多个分类数据集的数据特征,获得清洗数据类型、清洗手段以及清洗顺序;

[0025] 根据清洗数据类型、清洗手段以及清洗顺序,确定每个分类数据集的清洗规则。

[0026] 在第一方面的一种可能实现方式中,对超声波流量计上传的原始数据包进行拆分,获得多个分类数据集,包括:

[0027] 对超声波流量计上传的原始数据包进行拆解,获得多源数据;

[0028] 对多源数据进行分类,获得多个分类数据集。

[0029] 第二方面,本申请实施例提供一种基于物联网的超声波流量计数据清洗系统,应用于物联网系统,物联网系统包括:依次交互的用户平台、服务平台、管理平台、传感网络平台以及对象平台,对象平台用于接入超声波流量计,管理平台包括:

[0030] 拆分模块,拆分模块用于对超声波流量计上传的原始数据包进行拆分,获得多个分类数据集;

[0031] 确定模块,确定模块用于根据多个分类数据集的数据特征,确定每个分类数据集的清洗规则;

[0032] 清洗模块,清洗模块用于根据清洗规则,对每个分类数据集分别进行数据清洗,获得清洗数据集;

[0033] 整合模块,整合模块用于根据原始数据包中的数据分布信息,对清洗数据集进行整合,获得目标数据包。

[0034] 第三方面,本申请实施例提供一种计算机可读存储介质,储存有计算机程序,计算机程序被处理器加载执行时,实现如上述第一方面中任一项提供的基于物联网的超声波流量计数据清洗方法。

[0035] 第四方面,本申请实施例提供一种基于物联网的超声波流量计数据清洗装置,包括处理器及存储器,其中,

[0036] 存储器用于存储计算机程序;

[0037] 处理器用于加载执行计算机程序,以使基于物联网的超声波流量计数据清洗装置执行如上述第一方面中任一项提供的基于物联网的超声波流量计数据清洗方法。

[0038] 与现有技术相比,本申请的有益效果是:

[0039] 本申请实施例提出的一种基于物联网的超声波流量计数据清洗方法、系统及装

置,该方法包括:对超声波流量计上传的原始数据包进行拆分,获得多个分类数据集;根据多个分类数据集的数据特征,确定每个分类数据集的清洗规则;根据清洗规则,对每个分类数据集分别进行数据清洗,获得清洗数据集;根据原始数据包中的数据分布信息,对清洗数据集进行整合,获得目标数据包。本申请通过对象平台将超声波流量计接入物联网系统,实现对超声波流量计所采集的多源数据的处理,首先将超声波流量计上传的数据包进行拆分,得到其中多源数据的类别,并将其分类得到分类数据集,然后根据分类数据集的数据特征,也即多源数据的数据特征,分别制定适合不同类别数据的清洗规则,分别进行数据清洗后,采集数据的质量得到提升,最后根据原始数据包中的数据分布信息,将清洗后的数据集整合为数据包,以便于再传输的质量,并且整合依据原始数据包中的数据分布信息,在目标数据包中还原真实数据的分布状态,能为后处理分析提供质量更好的数据,有效提升后处理分析的水平。

附图说明

[0040] 图1为本申请实施例涉及的硬件运行环境的基于物联网的超声波流量计数据清洗装置结构示意图;

[0041] 图2为本申请实施例提供的基于物联网的超声波流量计数据清洗方法的流程图;

[0042] 图3为本申请实施例提供的基于物联网的超声波流量计数据清洗方法中物联网系统的框架示意图;

[0043] 图4为本申请实施例提供的基于物联网的超声波流量计数据清洗方法在一种实施方式下的逻辑框图;

[0044] 图中标记:101-处理器,102-通信总线,103-网络接口,104-用户接口,105-存储器。

具体实施方式

[0045] 应当理解,此处所描述的具体实施例仅仅用以解释本申请,并不用于限定本申请。

[0046] 参照附图1,附图1为本申请实施例方案涉及的硬件运行环境的基于物联网的超声波流量计数据清洗装置结构示意图,该电子设备可以包括:处理器101,例如中央处理器(Central Processing Unit,CPU),通信总线102、用户接口104,网络接口103,存储器105。其中,通信总线102用于实现这些组件之间的连接通信。用户接口104可以包括显示屏(Display)、输入单元比如键盘(Keyboard),可选用户接口104还可以包括标准的有线接口、无线接口。网络接口103可选的可以包括标准的有线接口、无线接口(如无线保真(Wireless-Fidelity,WI-FI)接口)。存储器105可选的可以是独立于前述处理器101的存储装置,存储器105可能是高速的随机存取存储器(Random Access Memory,RAM)存储器,也可能是稳定的非易失性存储器(Non-Volatile Memory,NVM),例如至少一个磁盘存储器;处理器101可以是通用处理器,包括中央处理器、网络处理器等,还可以是数字信号处理器、专用集成电路、现场可编程门阵列或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。

[0047] 本领域技术人员可以理解,附图1中示出的结构并不构成对基于物联网的超声波

流量计数据清洗装置的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件布置。

[0048] 如附图1所示,作为一种存储介质的存储器105中可以包括操作系统、网络通信模块、用户接口模块以及基于物联网的超声波流量计数据清洗系统。

[0049] 在附图1所示的基于物联网的超声波流量计数据清洗装置中,网络接口103主要用于与网络服务器进行数据通信;用户接口104主要用于与用户进行数据交互;本申请中的处理器101、存储器105可以设置在基于物联网的超声波流量计数据清洗装置中,基于物联网的超声波流量计数据清洗装置通过处理器101调用存储器105中存储的基于物联网的超声波流量计数据清洗系统,并执行本申请实施例提供的基于物联网的超声波流量计数据清洗方法。

[0050] 参照附图2,基于前述实施例的硬件设备,本申请的实施例提供一种基于物联网的超声波流量计数据清洗方法,应用于物联网系统,物联网系统包括:依次交互的用户平台、服务平台、管理平台、传感网络平台以及对象平台,对象平台用于接入超声波流量计,基于物联网的超声波流量计数据清洗方法包括以下步骤:

[0051] S10:对超声波流量计上传的原始数据包进行拆分,获得多个分类数据集。

[0052] 在具体实施过程中,超声波流量计是指一种基于超声波在流动介质中传播速度等于被测介质的平均流速与声波在静止介质中速度的矢量和的原理开发的流量计。单一设备就能实现流量、流速、压力、压强等多方面的数据,因此,本申请中所说的多源数据,并非指不同设备测量,而是同一超声波流量计测量的不同类别的数据,被测的对象可以是输水管路、输气管路、输油管路等。为了便于上传数据至物联网平台,会在超声波流量计一端就将多源数据进行封装、压缩等处理,打包为数据包的形式以便于保障传输性能。具体来说:对超声波流量计上传的原始数据包进行拆分,获得多个分类数据集,包括:

[0053] 对超声波流量计上传的原始数据包进行拆解,获得多源数据;

[0054] 对多源数据进行分类,获得多个分类数据集。

[0055] 在具体实施过程中,首先将数据包进行拆解,完全展开所有采集的多源数据,然后按照超声波流量计采集的类别进行分类,如数据属于流量数据、流速数据或者压强数据等,为了实现更准确的分类,在采集的时候可以对数据打上不同类别的标签,而后根据标签即可快速准确完成分类数据集的获取。

[0056] S20:根据多个分类数据集的数据特征,确定每个分类数据集的清洗规则。

[0057] 在具体实施过程中,分类数据集的数据特征,也就是多源数据的数据特征,比如数据量、数据的离散程度、数据的连续性、可变性等,不同源的数据的数据特征是不同的,虽然通用的数据清洗方式能够在一定程度上完成工作,但为了提升数据的质量,需要根据数据的特征来为分类数据集分别制定清洗规则。具体来说:根据多个分类数据集的数据特征,确定每个分类数据集的清洗规则,包括:

[0058] 根据分类数据集的数据特征,获得清洗数据类型、清洗手段以及清洗顺序;

[0059] 根据清洗数据类型、清洗手段以及清洗顺序,确定每个分类数据集的清洗规则。

[0060] 在具体实施过程中,清洗规则为指导数据清洗进行的规则,包括了清洗数据的顺序以及手段等多种信息。首先需要确认清洗数据类型,比如数据类型为图像数据、文本数据、数值数据,或者根据数据基本类型划分为整数类型、浮点类型等;然后确认清洗手段,可

包括丢弃、补全、转换、筛选、聚类回归等,比如丢弃可以是直接删除某些孤岛数据或者离散程度较大的数据,或者是删除某些重复数据;补全则是对缺失数据的补充,可以分为原本有数据缺失的补全,和原本无数据缺失的补全。确认前述信息后,在按照数据清洗的要求或经验对清洗顺序进行规划,最终得到针对不同分类数据集特点的清洗规则,随后按照清洗规则执行数据清洗任务即可,即:

[0061] S30:根据清洗规则,对每个分类数据集分别进行数据清洗,获得清洗数据集。

[0062] S40:根据原始数据包中的数据分布信息,对清洗数据集进行整合,获得目标数据包。

[0063] 在具体实施过程中,原始数据包中虽然含有一些缺陷需要进行数据清洗来提升质量,但是原始数据包才包含有与真实情况最接近的数据分布,比如在组合分析中,为了分析流量与压强在实际使用中的影响,二者各自的数据分布以及在整个数据中的分布情况就十分重要。因此对其进行整合,来实现真实数据分布的还原。具体来说:根据原始数据包中的数据分布信息,对清洗数据集进行整合,获得目标数据包,包括:

[0064] 根据原始数据包中的数据分布信息,获得清洗数据集的第一数据分布信息以及每个清洗数据集中数据的第二数据分布信息;

[0065] 根据第一数据分布信息,对清洗数据集进行一次整合,获得整合数据;

[0066] 根据第二数据分布信息,对整合数据进行二次整合,获得目标数据包。

[0067] 在具体实施过程中,由于原始数据包中各多源数据有不同的分布,多源数据自身包括的所有同源数据也有不同的分布,尤其是在超声波流量计的使用中,通常是会连续不断地采集数据,以时间序列排布多个数据,因此先考虑各个多源数据也即清洗数据集之间的分布,也即第一数据分布信息,第一次进行整合,将不同类别的数据分布还原,然后再根据第二数据分布信息,对每个类别中的数据进行二次整合,得到目标数据包。

[0068] 在一种实施例中,根据第二数据分布信息,对整合数据进行二次整合,获得目标数据包,包括:

[0069] 根据第二数据分布信息,获得每个清洗数据集中数据的时间序列排布;

[0070] 根据时间序列排布对清洗数据集中数据进行二次整合,以使整合数据中各清洗数据集的序列长度的比值等于各分类数据集的序列长度的比值,获得目标数据包。

[0071] 在具体实施过程中,由于时间序列相对各个数据集的内部数据更为重要,因此对这部分数据的分布还原作进一步设计,将待处理的数据集,也即清洗数据集中的数据按照时间序列进行排布,由于进行了数据清洗之后的数据,相对原始数据的数据量、序列都会出现变化,基于时间序列采集的数据并不能完全对应,因此为了还原真实数据的分布,就在保留已有数据的基础上进行同比例的调整,确保原始数据和目标数据中,多源数据占比保持相同。

[0072] 本实施例中,通过对象平台将超声波流量计接入物联网系统,实现对超声波流量计所采集的多源数据的处理,首先将超声波流量计上传的数据包进行拆分,得到其中多源数据的类别,并将其分类得到分类数据集,然后根据分类数据集的数据特征,也即多源数据的数据特征,分别制定适合不同类别数据的清洗规则,分别进行数据清洗后,采集数据的质量得到提升,最后根据原始数据包中的数据分布信息,将清洗后的数据集整合为数据包,以便于再传输的质量,并且整合依据原始数据包中的数据分布信息,在目标数据包中还原真

实数据的分布状态,能为后处理分析提供质量更好的数据,有效提升后处理分析的水平。

[0073] 在一种实施例中,根据时间序列排布对清洗数据集中数据进行二次整合,以使整合数据中各清洗数据集的序列长度的比值等于各分类数据集的序列长度的比值,获得目标数据包之后,方法还包括:

[0074] 判断目标数据包和原始数据包中数据的相似度是否小于相似度阈值;

[0075] 响应于相似度小于相似度阈值,返回对超声波流量计上传的原始数据包进行拆分,获得分类数据集的步骤,循环至相似度不小于相似度阈值。

[0076] 在具体实施过程中,为了验证整合处理后数据的质量,判断目标数据与原始数据之间的相似度,比如用曼哈顿距离、余弦相似度等衡量数据之间是否足够接近,若其相似度不小于阈值,说明目标数据保留了足够多的原始数据信息,若相似度小于阈值,那么可能是整个数据处理过程有异常,返回到拆分数据的步骤,重新进行分类、数据清洗、整合后再一次进行验证。

[0077] 考虑到时间序列相对数据的重要性,考虑将相似度分解为两部分,也即:判断目标数据包和原始数据包中数据的相似度是否小于相似度阈值之前,方法还包括:

[0078] 根据目标数据包和原始数据包,分别获得目标数据包和原始数据包中数据的分布相似度以及目标数据包和原始数据包中数据的序列相似度;

[0079] 向分布相似度和序列相似度分配不同权重占比,获得目标数据包和原始数据包中数据的相似度。

[0080] 相似度数据由分布相似度和序列相似度组合而成,在应用时根据实际要求对其权重占比进行分配,以公式表达为:

$$[0081] \quad E = E_1 * \varepsilon_1 + E_2 * \varepsilon_2$$

[0082] 其中, E 、 E_1 、 E_2 均在0-1之间, E 为相似度、 E_1 为分布相似度、 E_2 为序列相似度, ε_1 为分布相似度的权重占比, ε_2 为序列相似度的权重占比, $\varepsilon_1 + \varepsilon_2 = 1$,比如后处理分析需要研究随流量与流速的影响关系,更看重数据分布、数据量的关系,时序数据的影响偏小,那么可以设置 ε_1 偏大、 ε_2 偏小,如 $\varepsilon_1 = 0.9$ 、 $\varepsilon_2 = 0.1$;又如后处理分析需要研究随时间变化下,流量与压力的影响关系,在保有数据分布的同时,需要放大时序数据的影响,那么可以设置 $\varepsilon_1 = 0.7$ 、 $\varepsilon_2 = 0.3$ 。如上实施方式,将相似度数据分解后可进行控制调整,以使其与后处理分析工作关联,进一步提升数据的质量,提升后处理分析的水平。

[0083] 本申请实施例中所提供的物联网系统 in 应用场景下的框架如附图3所示,可包括依次交互的用户平台、服务平台、管理平台、传感网络平台以及对象平台,形成基本的五平台构架,其中,对象平台可包括超声波流量计对象分平台;传感网络平台可包括设备管理模块与数据传输管理模块,设备管理模块可包括网络管理单元、指令管理单元以及设备状态管理单元,数据传输管理模块可包括数据协议管理单元、数据解析单元、数据分类单元、数据传输监控单元以及数据传输安全单元;管理平台可包括设备管理分平台、业务管理分平台和数据中心,设备管理分平台和业务管理分平台可分别与数据中心进行交互,设备管理分平台可包括设备运行状态监控管理单元、计量数据监控管理单元、设备参数管理单元以及设备生命周期管理单元,业务管理分平台可包括营收管理单元、工商户管理单元、报装管理单元、消息管理单元、调度管理单元、购销差管理单元、运行分析管理单元以及综合业务管理单元,基于物联网的超声波流量计数据清洗系统的各功能模块可集成至管理平台,

并储存在数据中心中实现调用;服务平台可包括使用服务模块、运营服务模块以及安全服务模块;用户平台可包括普通用户模块、政府用户模块以及监管用户模块,通过上述五平台为基础的物联网系统的各个功能平台之间的交互,建立完善的闭环信息运行逻辑,确保了感知信息和控制信息的有序运行,实现设备本地操作安全的智慧化管理。

[0084] 参照附图4,在如附图4所示的实施方式下,对本申请做进一步说明:

[0085] 首先,由超声波流量计将采集到的多源数据形成原始数据包进行上传,物联网平台接收到该原始数据包后先进行拆分,再以多源数据的类型进行分类,得到若干个分类数据集;其次得到分类数据集的数据特征,从而针对不同的分类数据制定不同的清洗规则,以提升数据清洗的质量,进而按照清洗规则进行数据清洗得到清洗数据集,为了提供更高质量数据以供后处理分析使用,按照原始数据包中数据分布的特点对清洗后的数据进行还原,以将真实数据的分布状态尽可能保留在清洗后的数据中,整合过程中先以第一数据分布信息进行一次整合,将不同源数据在原始数据中的分布情况还原,然后再以第二数据分布信息对整合数据进行二次整合,还原多源数据各自包含的同源数据的分布情况,由此得到了在数据量、数据分布上都与真实数据保持高度相似的数据来支持后处理分析;最后在后处理分析之前验证整合还原的数据与真实数据的相似度,以确保数据在清洗、整合过程中没有出现影响还原数据质量的情况,验证依据相似度进行,考虑到超声波流量计采集数据具有时序的特点,并结合实际后处理分析的需求,将目标数据包与原始数据包之间的相似度评价分解为分布相似度与序列相似度,一方面考虑数据量、数据分布的影响,一方面引入时间序列分布对数据自身产生的影响,调整两个相似度各自的权重占比,耦合得到与数据特点、后处理分析相适应的相似度,进而判断该相似度是否小于相似度阈值即可,若小于,则说明前述的操作中存在较大偏差,需要返回读取原始数据包,重新再按照前述步骤进行数据清洗、整合等操作,若不小于,说明目标数据包与原始数据包的数据在各个方面都高度一致,可以用于后处理分析,以实现后处理分析水平的提升。

[0086] 基于与前述实施例中同样的发明构思,本申请实施例还提供一种基于物联网的超声波流量计数据清洗系统,应用于物联网系统,物联网系统包括:依次交互的用户平台、服务平台、管理平台、传感网络平台以及对象平台,对象平台用于接入超声波流量计,管理平台包括:

[0087] 拆分模块,拆分模块用于对超声波流量计上传的原始数据包进行拆分,获得分类数据集;

[0088] 确定模块,确定模块用于根据分类数据集的数据特征,确定每个分类数据集的清洗规则;

[0089] 清洗模块,清洗模块用于根据清洗规则,对分类数据集分别进行数据清洗,获得清洗数据集;

[0090] 整合模块,整合模块用于根据原始数据包中的数据分布信息,对清洗数据集进行整合,获得目标数据包。

[0091] 本领域技术人员应当理解,实施例中的各个模块的划分仅仅是一种逻辑功能的划分,实际应用时可以全部或部分集成到一个或多个实际载体上,且这些模块可以全部以软件通过处理单元调用的形式实现,也可以全部以硬件的形式实现,或是以软件、硬件结合的形式实现,需要说明的是,本实施例中基于物联网的超声波流量计数据清洗系统中各模块

是与前述实施例中的基于物联网的超声波流量计数据清洗方法中的各步骤一一对应,因此,本实施例的具体实施方式可参照前述基于物联网的超声波流量计数据清洗方法的实施方式,这里不再赘述。

[0092] 基于与前述实施例中同样的发明构思,本申请的实施例还提供一种计算机可读存储介质,储存有计算机程序,计算机程序被处理器加载执行时,实现如本申请实施例提供的基于物联网的超声波流量计数据清洗方法。

[0093] 基于与前述实施例中同样的发明构思,本申请的实施例还提供一种基于物联网的超声波流量计数据清洗装置,包括处理器及存储器,其中,

[0094] 存储器用于存储计算机程序;

[0095] 处理器用于加载执行计算机程序,以使基于物联网的超声波流量计数据清洗装置执行如本申请实施例提供的基于物联网的超声波流量计数据清洗方法。

[0096] 在一些实施例中,计算机可读存储介质可以是FRAM、ROM、PROM、EPROM、EEPROM、闪存、磁表面存储器、光盘、或CD-ROM等存储器;也可以是包括上述存储器之一或任意组合的各种设备。计算机可以是包括智能终端和服务器在内的各种计算设备。

[0097] 在一些实施例中,可执行指令可以采用程序、软件、软件模块、脚本或代码的形式,按任意形式的编程语言(包括编译或解释语言,或者声明性或过程性语言)来编写,并且其可按任意形式部署,包括被部署为独立的程序或者被部署为模块、组件、子例程或者适合在计算环境中使用的其它单元。

[0098] 作为示例,可执行指令可以但不一定对应于文件系统中的文件,可以可被存储在保存其它程序或数据的文件的一部分,例如,存储在超文本标记语言(HTML,Hyper Text Markup Language)文档中的一个或多个脚本中,存储在专用于所讨论的程序的单个文件中,或者,存储在多个协同文件(例如,存储一个或多个模块、子程序或代码部分的文件)中。

[0099] 作为示例,可执行指令可被部署为在一个计算设备上执行,或者在位于一个地点的多个计算设备上执行,又或者,在分布在多个地点且通过通信网络互连的多个计算设备上执行。

[0100] 需要说明的是,在本文中,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者系统不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者系统所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、方法、物品或者系统中还存在另外的相同要素。

[0101] 上述本申请实施例序号仅仅为了描述,不代表实施例的优劣。

[0102] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到上述 实施例方法可借助软件加必需的通用硬件平台的方式来实现,当然也可以通 过硬件,但很多情况下前者是更佳的实施方式。基于这样的理解,本申请的 技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质(如只读存储器/随机存取存储器、磁碟、光 盘)中,包括若干指令用以使得一台多媒体终端设备(可以是手机,计算机,电视接收机,或者网络设备等)执行本申请各个实施例所述的方法。

[0103] 综上,本申请提供的一种基于物联网的超声波流量计数据清洗方法、系统及装置,

该方法包括:对超声波流量计上传的原始数据包进行拆分,获得多个分类数据集;根据多个分类数据集的数据特征,确定每个分类数据集的清洗规则;根据清洗规则,对每个分类数据集分别进行数据清洗,获得清洗数据集;根据原始数据包中的数据分布信息,对清洗数据集进行整合,获得目标数据包。本申请通过对象平台将超声波流量计接入物联网系统,实现对超声波流量计所采集的多源数据的处理,首先将超声波流量计上传的数据包进行拆分,得到其中多源数据的类别,并将其分类得到分类数据集,然后根据分类数据集的数据特征,也即多源数据的数据特征,分别制定适合不同类别数据的清洗规则,分别进行数据清洗后,采集数据的质量得到提升,最后根据原始数据包中的数据分布信息,将清洗后的数据集整合为数据包,以便于再传输的质量,并且整合依据原始数据包中的数据分布信息,在目标数据包中还原真实数据的分布状态,能为后处理分析提供质量更好的数据,有效提升后处理分析的水平。

[0104] 以上所述仅为本申请的较佳实施例,并不用以限制本申请,凡在本申请的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本申请的保护范围之内。

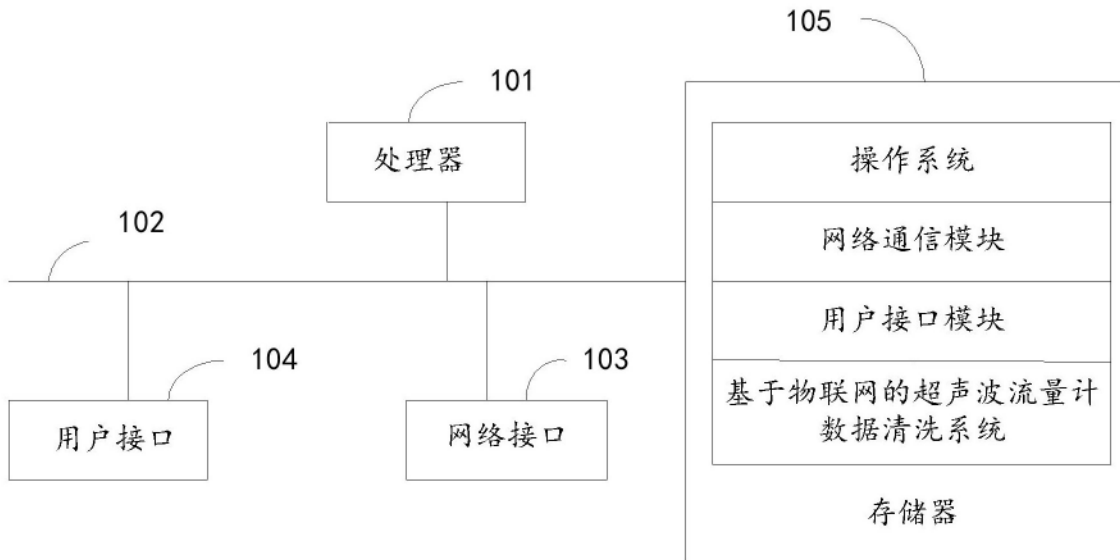


图1

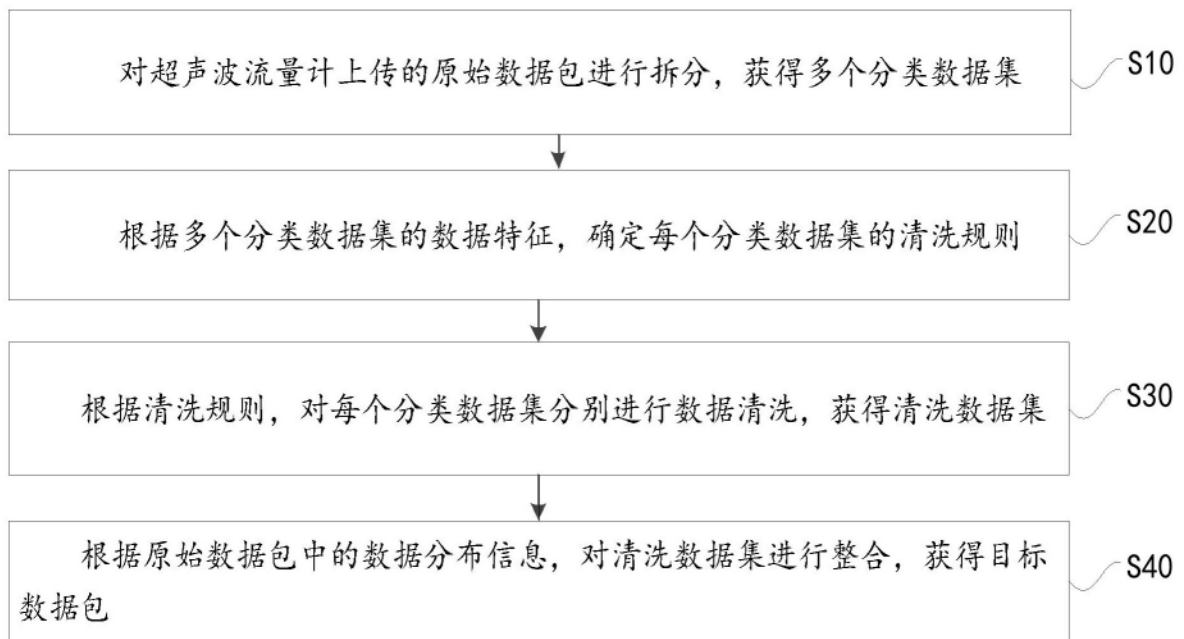


图2

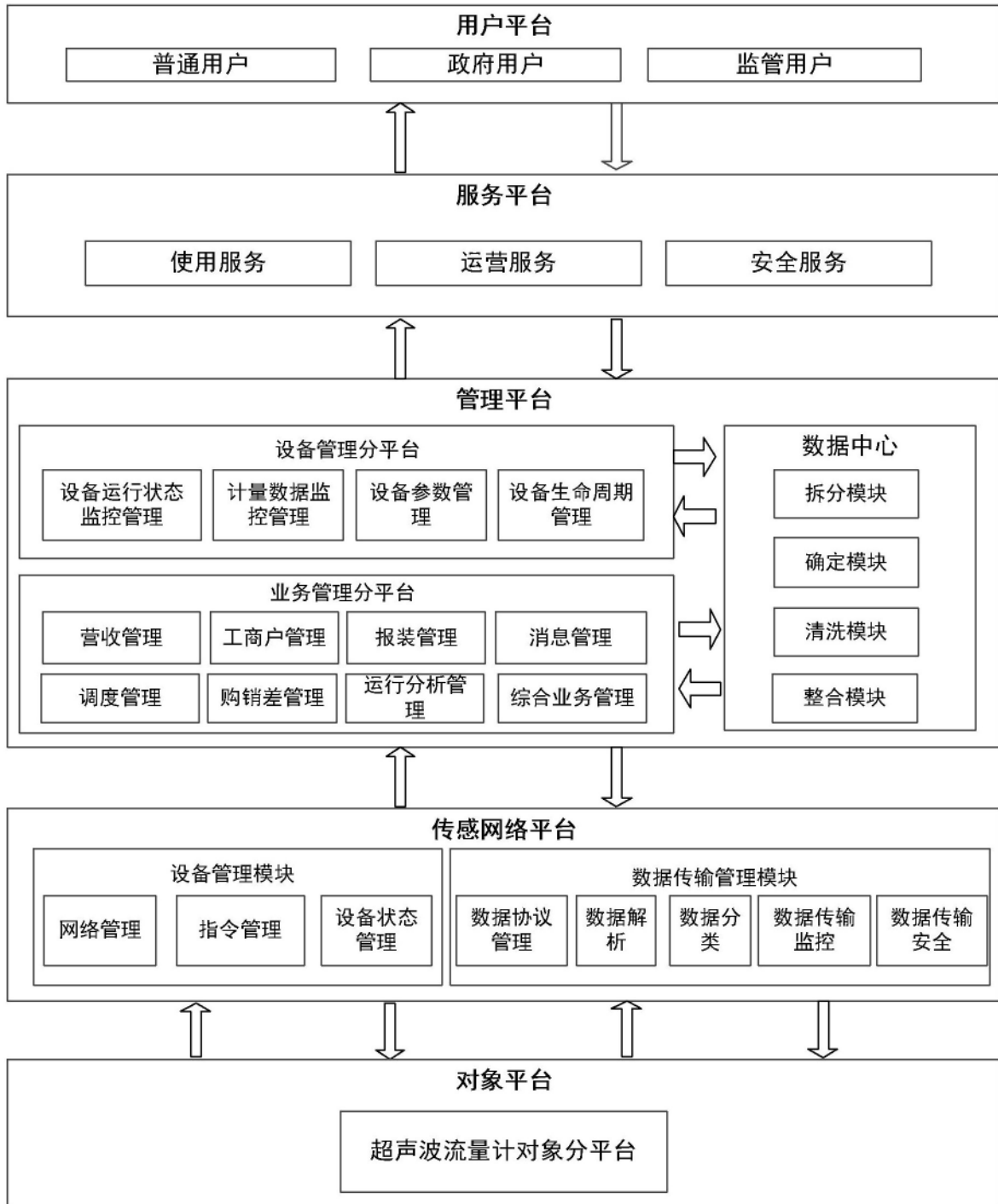


图3

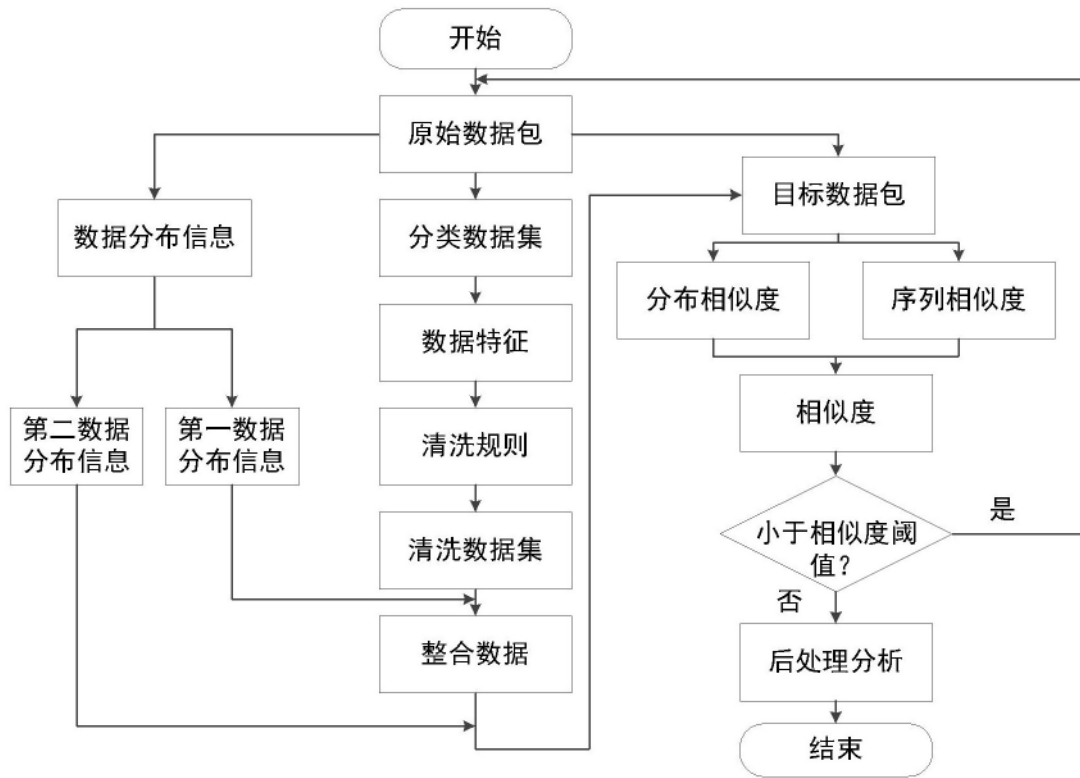


图4