



(12) 发明专利

(10) 授权公告号 CN 107256219 B

(45) 授权公告日 2021.05.14

(21) 申请号 201710270005.X

(51) Int.Cl.

(22) 申请日 2017.04.24

G06F 16/22 (2019.01)

(65) 同一申请的已公布的文献号

G06F 16/25 (2019.01)

申请公布号 CN 107256219 A

G06F 16/27 (2019.01)

(43) 申请公布日 2017.10.17

G06F 16/26 (2019.01)

(73) 专利权人 卡斯柯信号有限公司

G06F 16/24 (2019.01)

地址 200070 上海市闸北区天目中路428号

G06N 5/04 (2006.01)

凯旋门大厦27层C/D室

B61L 27/00 (2006.01)

专利权人 上海申通地铁集团有限公司

H04L 12/24 (2006.01)

审查员 张培

(72) 发明人 连爱娥 李文涛 胡恩华 魏盛昕

刘晓峰 周星宇 余亮星

(74) 专利代理机构 上海科盛知识产权代理有限公司 31225

代理人 应小波

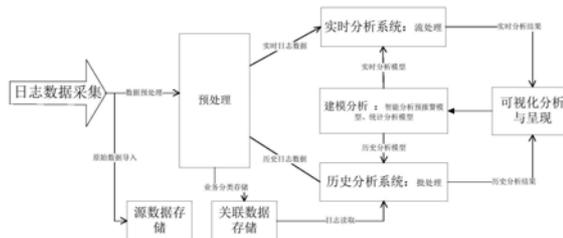
权利要求书2页 说明书5页 附图2页

(54) 发明名称

应用于自动列车控制系统海量日志的大数据融合分析方法

(57) 摘要

本发明涉及一种应用于自动列车控制系统海量日志的大数据融合分析方法,其包括以下步骤:(1)对系统日志中的业务可分析项进行基础数据类型定义;(2)对系统融合分析业务进行建模;(3)基于开放、标准原则实现统一的日志采集过程;(4)基于融合分析数据处理规则实现日志数据的预处理和存储;(5)基于业务分析模型实现跨系统日志关联分析;(6)通过统一接口实现日志分析结果的可视化展示。与现有技术相比,本发明具有及时诊断系统间的异常,可有效地减少维护人员的工作量等优点。



1. 一种应用于自动列车控制系统海量日志的大数据融合分析方法,其特征在于,包括以下步骤:

(1) 对系统日志中的业务可分析项进行基础数据类型定义:从系统日志数据中抽象出统一的数据类型定义,完整定义日志数据的基础分类方法及内容,分类结果作为后续分析的基础;

(2) 对系统融合分析业务进行建模:通过业务维度抽取子系统中关联的数据定义类型并建立逻辑模型;

(3) 基于开放、标准原则实现统一的日志采集过程:基于开放、标准原则提取并建立统一的日志采集框架封装,实现实时日志和历史日志的融合,实现内部的一致性表示;

(4) 基于融合分析数据处理规则实现日志数据的预处理和存储:根据业务规则及日志数据特性,设计数据处理规则,通过动态策略解析并提取日志信息,通过单业务维度或多业务维度实现数据的按需存储;

(5) 基于业务分析模型实现跨系统日志关联分析:实现基于业务维度的分析算法模型的封装实现,应用于实时分析和历史分析,实现实时数据和历史数据可扩展性的同步分析;

(6) 通过统一接口实现日志分析结果的可视化展示:为数据共享服务层提供标准化的统一接口;

所述的步骤(1)中的统一的数据类型定义具体如下:

建立基础数据类型定义表,通过该表按层次定义系统基础数据,如下:

(11) 第一层级为基本类型,包括开关量、多状态量、模拟量、字符量、曲线、文本、二进制;

(12) 第二层级为具体的类型码关联具体的业务类型;

(13) 将上述分类关系整理为具备唯一标识表示特定业务类型的数据表形式,供后续分析统一进行引用;

所述的(4)中的基于融合分析数据处理规则实现日志数据的预处理和存储具体为:

(41) 数据预处理:对系统日志进行有效整合、清洗、转换和提取:

(411) 预处理是根据业务规则预设了处理规范及参数,对日志采集模块采集来的日志进行特征提取,对多源异构提取的相同特征进行特征融合,以实现可观的信息压缩,形成结构化更强的数据,信息压缩后的结构化数据建立了信号设备与特征数据的映射关系;

(412) 对于无法拆解的日志信息暂时以字符串类型进行存储,用于后续知识库丰富后进行结构化和非结构化数据的相互关联挖掘分析;

(413) 子系统数据的融合:通过合并时间数据、事件数据,或者通过多个数据间共有的关键字来合并,或者根据自定义的条件进行合并;

(42) 数据存储:依据数据处理规则,将同类业务维度的信息统一以数据库和数据表的方式实现结构化数据管理,以数据库作为命名空间,其中存储方式包括:

(421) 对业务数据进行划分,同一类型的数据作为一张数据表进行存储,数据表按照时间字段进行分区;

(422) 根据业务维度的动态存储:按分析模型建立设备维度,车次维度的数据表,按时间进行分区存储;

所述的(6)中的通过统一接口实现日志分析结果的可视化展示具体为:

基于自动列车控制系统关键设备和关键部件的业务逻辑,提供一个可视化的界面,接收用户的请求,用户通过界面把任务命令传输到系统中,等待系统返回接口数据:

(61) 实时的监测指标数据、实时预报警信息前台展示;

(62) 系统交互分析:通过可视界面向系统发送数据采集预处理归约参数、向分析系统发送建模数据及其他的交互信息;

(63) 根据不同数据类型提供通用的固定组合查询统计模式,包括模拟量提供统计设定时间内超限数据的时间或次数统计功能;开关量和多状态量提供设定时间内的满足条件的次数和时间,结果提供图表展示功能;

(64) 提供多列多变量的组合逻辑查询:车载日志需展示列车停不准统计分析、列车紧急制动统计分析和列车位置信息查询展示;

采用基于业务分析模型的统一建模方法实现跨系统的日志融合分析,对自动列车控制系统关键设备和关键部件的运行状态进行在线监视,对车载系统和地面系统间的过程状态和环节进行关联分析;

所述的步骤(3)中基于开放、标准的原则实现统一的日志采集过程具体为:

日志采集将实时的日志信息导入到系统中并将已经存在的日志信息导入到系统中,采集框架实现统一的日志采集过程:

(31) 为日志源头设置代理监控,包括文本日志、关系数据库日志,对这些日志的路径及代理检测时间频率进行配置;

(32) 当监控到日志流后,以设定的格式发送给物理通道,该通道将收集到的日志流组成有序的序列;

(33) 接收器通过流的方式,按照设定的配置规则,将流写入分布式存储;

所述的(5)中的基于业务分析模型实现跨系统日志关联分析具体为:

日志关联分析分为实时数据分析和历史数据分析,

(51) 在自动列控控制系统中,实时数据分析用于故障诊断分析:选取系统预处理后的数据和人机界面或系统的预设参数按分析模型算法进行信息融合分析,在推理机作用下,调用知识库中的已有知识,进行逻辑推理,向用户提供故障诊断决策;在基于智能分析的故障诊断系统中加入智能学习模块,故障决策经智能学习模块反馈给知识库,并对相应的规约参数进行修改,不断更新及扩充知识库;

(52) 在自动列控控制系统中,历史数据分析是对设备全生命周期的数据进行统计分析,向用户提供设备状态的数据图及趋势报表信息;对设备参量进行预测分析,提供中远期的预测分析结果。

2. 根据权利要求1所述的一种应用于自动列车控制系统海量日志的大数据融合分析方法,其特征在于,所述的具体类型码关联具体的业务类型具体为:开关量进一步细分为紧急制动状态开关量、车辆空转开关量;模拟量进一步细分为轮径模拟量、牵引力模拟量。

3. 根据权利要求1所述的一种应用于自动列车控制系统海量日志的大数据融合分析方法,其特征在于,所述的步骤(2)对系统融合分析业务进行建模具体为:

基于统一的基础数据类型定义,对融合分析场景进行建模,对算法进行封装,分别为:

(21) 对于故障诊断业务,建立智能分析的故障树分析模型;

(22) 对于统计分析业务,建立趋势分析和预测模型。

应用于自动列车控制系统海量日志的大数据融合分析方法

技术领域

[0001] 本发明涉及大数据分析领域,尤其是涉及一种应用于自动列车控制系统海量日志的大数据融合分析方法。

背景技术

[0002] 当前,随着用户在故障突发时对系统日志的依赖程度越来越高,同时对于历史日志数据的统计预测分析要求逐步提升,急需一种有效的日志大数据融合分析方法。

[0003] 然而在目前的自动列控控制系统中,只是简单地把各子系统日志按协议解析后写入本地磁盘或同步到关系型数据库,以备后续的检索需要。这种分析方法存在的弊端:

[0004] 第一、数据范围限定于传统手段捕捉到的结构化数据,且只停留于对独立系统中可见的小规模的抽样数据进行分析,对于海量的非结构化或半结构化全量数据缺乏分析;

[0005] 第二、分析是建立在事先创建的关系数据模型之上的,缺乏对整体日志的综合分析,无法对跨系统日志间的关联关系进行深度挖掘利用;

[0006] 第三、无法满足故障预测时效性的要求,以上的分析方法是基于定向的批处理技术,在获取所需的关键数据之前无法进行预测分析。

[0007] 第四、无法满足扩展性需求,随着系统日志的海量增长,以上方法只能通过昂贵的硬件来满足其存储和处理能力需求,如大规模并行处理(MPP)系统或对称多处理(SMP)系统等。

[0008] 另外,随着系统扩展,各子系统日志格式及日志形式越来越趋向于多样化,分析复杂度已远超出了基本的分析系统和分析人员的能力,随其带来的可维护性问题也是瓶颈。

发明内容

[0009] 本发明的目的就是为了解决上述现有技术存在的缺陷而提供一种应用于自动列车控制系统海量日志的大数据融合分析方法。

[0010] 本发明的目的可以通过以下技术方案来实现:

[0011] 一种应用于自动列车控制系统海量日志的大数据融合分析方法,包括以下步骤:

[0012] (1)对系统日志中的业务可分析项进行基础数据类型定义:从系统日志数据中抽象出统一的数据类型定义,完整定义日志数据的基础分类方法及内容,分类结果作为后续分析的基础;

[0013] (2)对系统融合分析业务进行建模:通过业务维度抽取子系统中关联的数据定义类型并建立逻辑模型;其中业务维度包括时间维度、车次维度、设备维度;

[0014] (3)基于开放、标准原则实现统一的日志采集过程:基于开放、标准原则提取并建立统一的日志采集框架封装,实现实时日志和历史日志的融合,实现内部的一致性表示;其中日志包括但不限于列控车载系统日志、列控地面系统日志、联锁系统日志、轨旁系统日志;

[0015] 所述的日志为多源异构数据,采集过程包括两类:

[0016] 一类是基于流的增量实时采集:TCP socket、Web Service、Streaming;

[0017] 一类是基于历史数据的定时日志采集:File、DB、FTP。

[0018] (4) 基于融合分析数据处理规则实现日志数据的预处理和存储:根据业务规则及日志数据特性,设计数据处理规则,通过动态策略解析并提取日志信息,通过单业务维度或多业务维度实现数据的按需存储;

[0019] (5) 基于业务分析模型实现跨系统日志关联分析:实现基于业务维度的分析算法模型的封装实现,应用于实时分析和历史分析,实现实时数据和历史数据可扩展性的同步分析;

[0020] (6) 通过统一接口实现日志分析结果的可视化展示:为数据共享服务层提供标准化的统一接口。所述的数据共享服务层包括Restful API、Tcp Socket、Relation DB、NoSQL DB。

[0021] 所述的步骤(1)中的统一的数据类型定义具体如下:

[0022] 建立基础数据类型定义表,通过该表按层次定义系统基础数据,如下:

[0023] (11) 第一层级为基本类型,包括开关量、多状态量、模拟量、字符量、曲线、文本、二进制;

[0024] (12) 第二层级为具体的类型码关联具体的业务类型;

[0025] (13) 将上述分类关系整理为具备唯一标识表示特定业务类型的数据表形式,供后续分析统一进行引用。

[0026] 所述的具体的类型码关联具体的业务类型具体为:开关量进一步细分为紧急制动状态开关量、车辆空转开关量;模拟量进一步细分为轮径模拟量、牵引力模拟量。

[0027] 所述的步骤(2)对系统融合分析业务进行建模具体为:

[0028] 基于统一的基础数据类型定义,对融合分析场景进行建模,对算法进行封装,分别为:

[0029] (21) 对于故障诊断业务,建立智能分析的故障树分析模型;

[0030] (22) 对于统计分析业务,建立趋势分析和预测模型;

[0031] 所述的步骤(3)中基于开放、标准的原则实现统一的日志采集过程具体为:

[0032] 日志采集将实时的日志信息导入到系统中和将已经存在的日志信息导入到系统中,采集框架实现统一的日志采集过程:

[0033] (31) 为日志源头设置代理监控,包括文本日志、关系数据库日志,对这些日志的路径及代理检测时间频率进行配置;

[0034] (32) 当监控到日志流后,以设定的格式发送给物理通道,该通道将收集到的日志流组成有序的序列;

[0035] (33) 接收器通过流的方式,按照设定的配置规则,将流写入分布式存储。

[0036] 所述的(4)中的基于融合分析数据处理规则实现日志数据的预处理和存储具体为:

[0037] (41) 数据预处理:对系统日志进行有效整合、清洗、转换和提取;

[0038] (411) 预处理是根据业务规则预设了处理规范及参数,对日志采集模块采集来的日志进行特征提取,对多源异构提取的相同特征进行特征融合,以实现可观的信息压缩,形成结构化更强的数据,信息压缩后的结构化数据建立了信号设备与特征数据的映射关系;

[0039] (412) 对于无法拆解的日志信息暂时以字符串类型进行存储,用于后续知识库丰

富后进行结构化和非结构化数据的相互关联挖掘分析；

[0040] (413) 子系统数据的融合：通过合并时间数据、事件数据，或者通过多个数据间共有的关键字来合并，或者根据自定义的条件进行合并；

[0041] (42) 数据存储：依据数据处理规则，将同类业务维度的信息统一以数据库和数据表的方式实现结构化数据管理，以数据库作为命名空间，其中存储方式包括：

[0042] (421) 对业务数据进行划分，同一类型的数据作为一张数据表进行存储，数据表按照时间字段进行分区；

[0043] (422) 根据业务维度的动态存储：按分析模型建立设备维度，车次维度的数据表，按时间进行分区存储。

[0044] 所述的(5)中的基于业务分析模型实现跨系统日志关联分析具体为：

[0045] 日志关联分析分为实时数据分析和历史数据分析，

[0046] (51) 在自动列控控制系统中，实时数据分析用于故障诊断分析：选取系统预处理后的数据和人机界面或系统的预设参数按分析模型算法进行信息融合分析，在推理机作用下，调用知识库中的已有知识，向用户提供故障诊断决策；在基于智能分析的故障诊断系统中加入智能学习模块，故障决策经智能学习模块反馈给知识库，并对相应的规约参数进行修改，不断更新及扩充知识库；

[0047] (52) 在自动列控控制系统中，历史数据分析是对设备全生命周期的数据进行统计分析，向用户提供设备状态的数据图及趋势报表信息；对设备参量进行预测分析，提供中远期的预测分析结果。

[0048] 所述的(6)中的通过统一接口实现日志分析结果的可视化展示具体为：

[0049] 基于自动列车控制系统关键设备和关键部件的业务逻辑，提供一个可视化的界面，接收用户的请求，用户通过界面把任务命令传输到系统中，等待系统返回接口数据：

[0050] (61) 实时的监测指标数据、实时预报警信息前台展示；

[0051] (62) 系统交互分析：通过可视界面向系统发送数据采集预处理归约参数、向分析系统发送建模数据及其他的交互信息；

[0052] (63) 根据不同数据类型提供通用的固定组合查询统计模式，包括模拟量提供统计设定时间内超限数据的时间或次数统计功能；开关量和多状态量提供设定时间内的满足条件的次数和时间，结果提供图表展示功能；

[0053] (64) 提供多列多变量的组合逻辑查询：车载日志需展示列车停不准统计分析、列车紧急制动统计分析和列车位置信息查询展示。

[0054] 与现有技术相比，本发明基于分布式计算框架和存储技术，采用基于数据处理规则的动态调度方法实现日志数据的预处理和存储；采用基于业务分析模型的统一建模方法实现跨系统的日志融合分析，对自动列车控制系统关键设备和关键部件的运行状态进行在线监视，尤其对车载系统和地面系统间的过程状态和环节进行关联分析，及时诊断系统间的异常，可有效地减少维护人员的工作量。

附图说明

[0055] 图1是本发明一种应用于自动列车控制系统海量日志的大数据融合分析方法中系统的整体架构图；

[0056] 图2是本发明一种应用于自动列车控制系统海量日志的大数据融合分析方法中系统中对于数据预处理的示意图；

[0057] 图3是本发明一种应用于自动列车控制系统海量日志的大数据融合分析方法中系统中对于日志关联分析的示意图；

[0058] 图4是本发明一种应用于自动列车控制系统海量日志的大数据融合分析方法中系统中对于子系统融合分析的案例示意图。

具体实施方式

[0059] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明的一部分实施例,而不是全部实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例,都应属于本发明保护的范围。

[0060] 如图1所示,本发明应用于自动列车控制系统海量日志的大数据融合分析方法,包括以下步骤:

[0061] 1、对系统日志中的业务可分析项进行基础数据类型定义

[0062] 首先需要建立基础数据类型定义表。通过该表按层次定义系统基础数据,如下:

[0063] (1) 第一层级为基本类型,主要包括开关量、多状态量、模拟量、字符量、曲线、文本、二进制等;

[0064] (2) 第二层级为具体的类型码关联具体的业务类型,如开关量可进一步细分为紧急制动状态开关量、车辆空转开关量;模拟量可进一步细分为轮径模拟量、牵引力模拟量等;

[0065] (3) 将上述分类关系整理为具备唯一标识表示特定业务类型的数据表形式,供后续分析统一进行引用。

[0066] 2、对系统融合分析业务进行建模

[0067] 基于统一的基础数据类型定义,对融合分析场景进行建模,对算法进行封装。

[0068] (1) 对于故障诊断业务,建立智能分析的故障树分析模型;

[0069] (2) 对于统计分析业务,建立趋势分析和预测模型;

[0070] 3、基于开放、标准的原则实现统一的日志采集过程

[0071] 日志采集主要是将实时的日志信息导入到系统中并将已经存在的日志信息导入到系统中。采集框架实现统一的日志采集过程:

[0072] (1) 为日志源头设置代理监控,如文本日志,关系数据库日志等,对这些日志的路径及代理检测时间频率进行配置;

[0073] (2) 当监控到日志流后,以一定的格式发送给物理通道,该通道将收集到的日志流组成有序的序列;

[0074] (3) 接收器通过流的方式,按照一定的配置规则,将流写入分布式存储。

[0075] 4、基于融合分析数据处理规则实现日志数据的预处理和存储,如图2所示。

[0076] (1) 数据预处理:对系统日志进行有效整合、清洗、转换和提取

[0077] (1.1) 预处理是根据业务规则预设了处理规范及参数(信号系统设备信息表及数据描述规范),对日志采集模块采集来的日志进行特征提取,对多源异构提取的相同特征进

行特征融合,以实现可观的信息压缩,形成结构化更强的数据,信息压缩后的结构化数据建立了信号设备与特征数据的映射关系;

[0078] (1.2) 对于无法拆解的日志信息暂时以字符串类型进行存储,用于后续知识库丰富后进行结构化和非结构化数据的相互关联挖掘分析;

[0079] (1.3) 子系统数据的融合:可通过合并时间数据、事件数据,也可通过多个数据间共有的关键字来合并,还可以根据自定义的条件进行合并。

[0080] (2) 数据存储:依据数据处理规则,将同类业务维度的信息统一以数据库和数据表的方式实现结构化数据管理,以数据库作为命名空间,存储方式包括:

[0081] (2.1) 对业务数据进行划分,同一类型的数据作为一张数据表进行存储,数据表按照时间等不同类型字段进行分区;

[0082] (2.2) 根据业务维度的动态存储:按分析模型建立设备维度,车次维度的数据表,按时间进行分区存储;

[0083] 5、基于业务分析模型实现跨系统日志关联分析,如图3和图4所示。

[0084] 日志关联分析主要分为实时数据分析和历史数据分析。

[0085] (1) 在自动列控控制系统中,实时数据分析主要用于故障诊断分析:主要是选取系统预处理后的数据和人机界面或系统的预设参数按分析模型算法进行信息融合分析,在推理机作用下,调用知识库中的已有知识,进行逻辑推理,向用户反馈故障诊断决策。在基于智能分析的故障诊断系统中加入智能学习模块,故障决策经自智能学习模块反馈给知识库,并对相应的规约参数进行修改,不断更新及扩充识库。

[0086] (2) 在自动列控控制系统中,历史数据分析主要是对设备全生命周期的数据进行统计分析,向用户提供设备状态的数据图及趋势报表信息;对设备参量进行预测分析,提供中远期的预测分析结果。

[0087] 6、通过统一接口实现日志分析结果的可视化展示

[0088] 基于自动列车控制系统关键设备和关键部件的业务逻辑,提供一个可视化的界面,接收用户的请求,用户通过界面把任务命令传输到系统中,等待系统返回接口数据。

[0089] (1) 实时的监测指标数据、实时预报警信息前台展示;

[0090] (2) 系统交互分析:通过可视界面向系统发送数据采集预处理归约参数、向分析系统发送建模数据及其他的交互信息;

[0091] (3) 根据不同数据类型提供通用的固定组合查询统计模式,如模拟量提供统计一段时间内超限数据的时间或次数统计功能;开关量和多状态量提供一段时间内的满足条件的次数和时间,结果提供图表展示功能。

[0092] (4) 提供多列多变量的组合逻辑查询:车载日志需展示列车停不准统计分析、列车紧急制动统计分析和列车位置信息查询展示等。

[0093] 以上所述,仅为本发明的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到各种等效的修改或替换,这些修改或替换都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应以权利要求的保护范围为准。

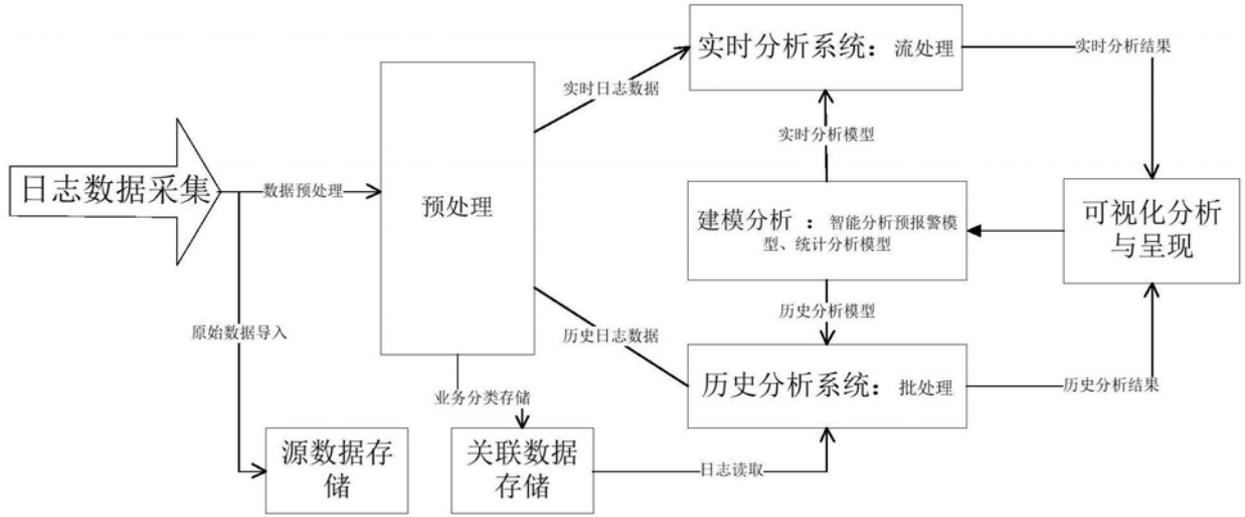


图1

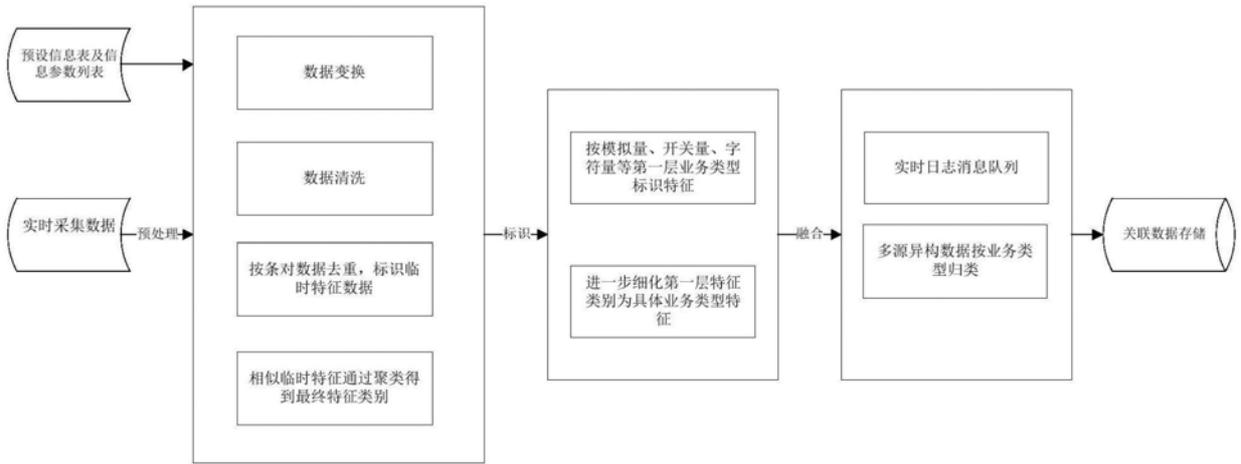


图2

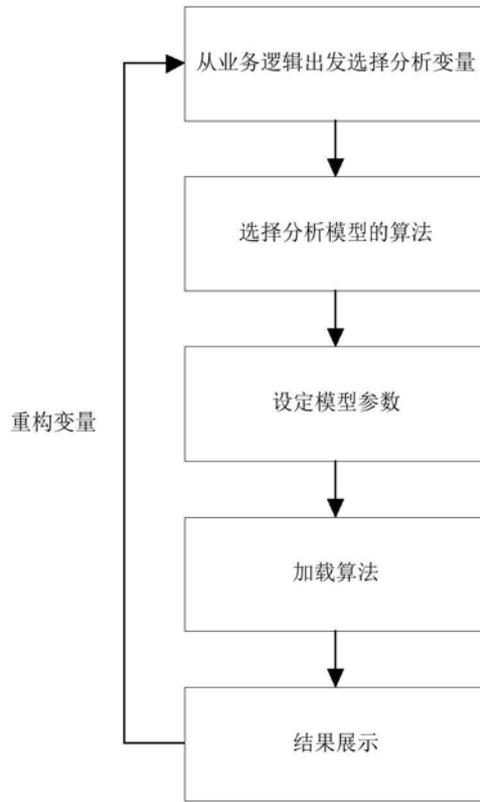


图3

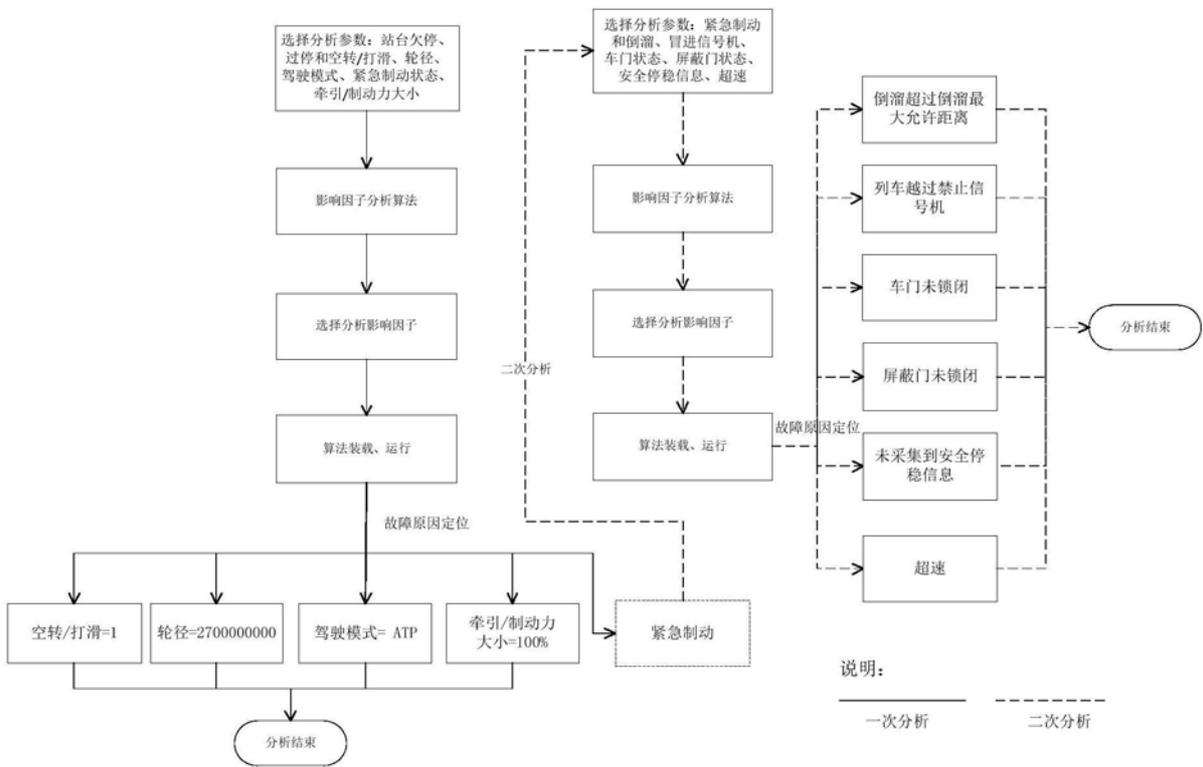


图4