

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 12 月 24 日 (24.12.2020)



(10) 国际公布号
WO 2020/253399 A1

(51) 国际专利分类号:
G06F 16/17 (2019.01)

(21) 国际申请号: PCT/CN2020/087835

(22) 国际申请日: 2020 年 4 月 29 日 (29.04.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910546292.1 2019 年 6 月 21 日 (21.06.2019) CN

(71) 申请人: 深圳前海微众银行股份有限公司
(**WEBANK CO., LTD**) [CN/CN]; 中国广东省深圳市南山区沙河西路 1819 号深圳湾科技生态园 7 栋 A 座, Guangdong 518052 (CN)。

(72) 发明人: 李悦 (**LI, Yue**); 中国广东省深圳市南山区沙河西路 1819 号深圳湾科技生态园 7 栋 A 座, Guangdong 518052 (CN)。曾可 (**ZENG, Ke**); 中国广东省深圳市南山区沙河西路 1819 号深圳湾科技生态园 7 栋 A 座, Guangdong 518052 (CN)。卢道和 (**LU, Daohe**); 中国广东省深圳

市南山区沙河西路 1819 号深圳湾科技生态园 7 栋 A 座, Guangdong 518052 (CN)。

(74) 代理人: 深圳市世纪恒程知识产权代理事务所 (**CENFO INTELLECTUAL PROPERTY AGENCY**); 中国广东省深圳市南山区西丽街道松坪山社区松坪山路 3 号奥特迅电力大厦 201, Guangdong 518052 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,

(54) **Title:** LOG CLASSIFICATION RULE GENERATION METHOD, DEVICE, APPARATUS, AND READABLE STORAGE MEDIUM

(54) 发明名称: 日志分类规则的生成方法、装置、设备及可读存储介质

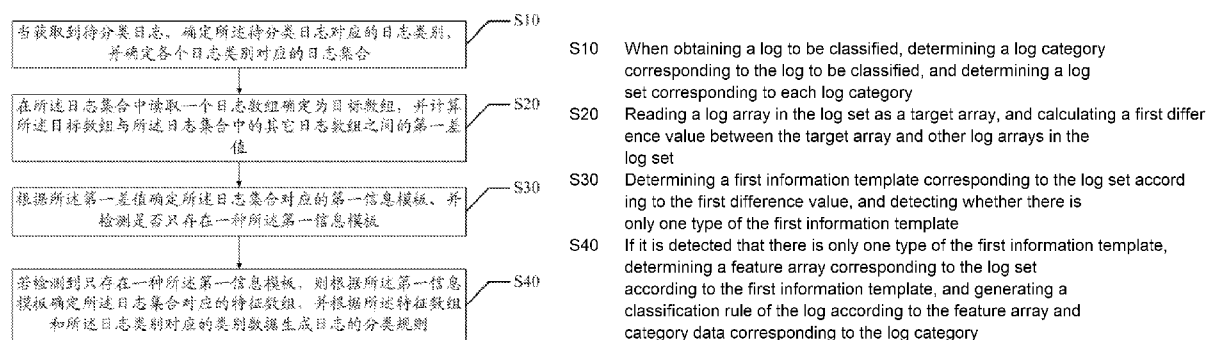


图 1

(57) **Abstract:** A log classification rule generation method, a device, an apparatus, and a readable storage medium, relating to the field of financial science and technology, wherein the method comprises the following steps: when obtaining a log to be classified, determining a log category corresponding to the log to be classified, and determining a log set corresponding to each log category (S10); reading a log array in the log set as a target array, and calculating a first difference value between the target array and other log arrays in the log set (S20); determining a first information template corresponding to the log set according to the first difference value, and detecting whether there is only one type of the first information template (S30); if it is detected that there is only one type of the first information template, determining a feature array corresponding to the log set according to the first information template, and generating a classification rule of the log according to the feature array and category data corresponding to the log category (S40).

NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布：

— 包括国际检索报告(条约第21条(3))。

(57) 摘要：一种日志分类规则的生成方法、装置、设备及可读存储介质，涉及金融科技领域，该方法包括步骤：当获取到待分类日志，确定待分类日志对应的日志类别，并确定各个日志类别对应的日志集合（S10）；在日志集合中读取一个日志数组确定为目标数组，并计算目标数组与日志集合中的其它日志数组之间的第一差值（S20）；根据第一差值确定日志集合对应的第一信息模板，并检测是否只存在一种第一信息模板（S30）；若检测到只存在一种第一信息模板，则根据第一信息模板确定日志集合对应的特征数组，并根据特征数组和日志类别对应的类别数据生成日志的分类规则（S40）。

日志分类规则的生成方法、装置、设备及可读存储介质

[0001] 本申请要求于2019年6月21日申请的、申请号为201910546292.1、名称为“日志分类规则的生成方法、装置、设备及可读存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

[0002] 本申请涉及金融科技（Fintech）的数据处理技术领域，尤其涉及一种日志分类规则的生成方法、装置、设备及可读存储介质。

背景技术

[0003] 随着金融科技，尤其是互联网科技金融的不断发展，越来越多的技术（如分布式、大数据、区块链Blockchain、人工智能等）应用在金融领域，但金融业也对技术提出了更高的要求，如在金融领域，当通过交换机、路由器和防火墙等实现线上金融交易时，需要对交易过程中产生的日志进行分类。

[0004] 在线上交易过程中，在交换机、路由器和防火墙等网络设备中会产生文本记录，该文本记录就是日志。各网络设备厂商依据内部标准，将交换机、路由器和防火墙等网络设备运行期间遇到的事件记为日志，一般每一行日志描述某个时间点某个对象最新的状态。因为描述每类事件的句子格式不一致，且同类事件在不同设备型号上的句子格式也可能不一致，所以日志是结构不规则的非结构化数据。非结构化数据是数据结构不规则或不完整，没有预定义的数据模型，不方便用数据库二维逻辑表来表现的数据，比如各类文本。

[0005] 目前常用的日志分类方法有如下两种，第一种是依据网络设备厂商和型号，人工分析网络设备的日志数据，然后定义每一类事件的正则表达式规则，由预先设置好的程序按照定义的正则表达式规则对日志进行分类。第二种是利用文本相似度算法计算每行日志样本之间的相似度，相似度大于设定值的为同类样本，利用同类样本生成分类所需的正则表达式规则，最后依据正则表达式规则对日志进行分类。第一种日志分类方法需要人工分析日志并定义分类所需的正则表达式，工作量大，日志分类成本高，正则表达式规则生成效率低下；第二种

日志分类方法的正则表达式来自日志样本，而分类依赖日志样本之间相似度的设定值，该方法只能通过设定值间接干预分类所需的正则表达式规则，生成正则表达式规则的过程繁琐。

[0006] 由此可知，目前生成日志分类所需的正则表达式规则过程繁琐，效率低下。

发明概述

技术问题

问题的解决方案

技术解决方案

[0007] 本申请的主要目的在于提供一种日志分类规则的生成方法、装置、设备及可读存储介质，旨在解决现有的生成日志分类所需的正则表达式规则过程繁琐，效率低下的技术问题。

[0008] 为实现上述目的，本申请提供一种日志分类规则的生成方法，所述日志分类规则的生成方法包括步骤：

[0009] 当获取到待分类日志，确定所述待分类日志对应的日志类别，并确定各个日志类别对应的日志集合；

[0010] 在所述日志集合中读取一个日志数组确定为目标数组，并计算所述目标数组与所述日志集合中的其它日志数组之间的第一差值；

[0011] 根据所述第一差值确定所述日志集合对应的第一信息模板，并检测是否只存在一种所述第一信息模板；

[0012] 若检测到只存在一种所述第一信息模板，则根据所述第一信息模板确定所述日志集合对应的特征数组，并根据所述特征数组和所述日志类别对应的类别数据生成日志的分类规则。

[0013] 此外，为实现上述目的，本申请还提供一种日志分类规则的生成装置，所述日志分类规则的生成装置包括：

[0014] 确定模块，用于当获取到待分类日志，确定所述待分类日志对应的日志类别，并确定各个日志类别对应的日志集合；

[0015] 读取模块，用于在所述日志集合中读取一个日志数组确定为目标数组；

[0016] 计算模块，用于计算所述目标数组与所述日志集合中的其它日志数组之间的第

一差值；

[0017] 所述确定模块还用于根据所述第一差值确定所述日志集合对应的第一信息模板；

[0018] 检测模块，用于检测是否只存在一种所述第一信息模板；

[0019] 所述确定模块还用于若检测到只存在一种所述第一信息模板，则根据所述第一信息模板确定所述日志集合对应的特征数组；

[0020] 生成模块，用于根据所述特征数组和所述日志类别对应的类别数据生成日志的分类规则。

[0021] 此外，为实现上述目的，本申请还提供一种日志分类规则的生成设备，所述日志分类规则的生成设备包括存储器、处理器和存储在所述存储器上并可在所述处理器上运行的日志分类规则的生成程序，所述日志分类规则的生成程序被所述处理器执行时实现如上所述的日志分类规则的生成方法的步骤。

[0022] 此外，为实现上述目的，本申请还提供一种计算机可读存储介质，所述计算机可读存储介质上存储有日志分类规则的生成程序，所述日志分类规则的生成程序被处理器执行时实现如上所述的日志分类规则的生成方法的步骤。

[0023] 本申请通过在获取到待分类日志后，确定待分类日志中各个日志类别对应的日志集合，在日志集合中读取一个日志数组确定为目标数组，根据目标数组与日志集合中的其它日志数组之间的第一差值确定日志集合对应的第一信息模板，若检测到只存在一种第一信息模板，则根据第一信息模板确定日志集合对应的特征数组，根据特征数组和日志类别对应的类别数据生成日志的分类规则。本实施例基于日志数组和模板的概念，得出特征数组，根据特征数组和日志类别生成日志的分类规则，通过模板和数组的直观性和可调节性便于用户快速调整分类规则，且不再需要生成正则表达式规则对日志进行分类，而是通过数组和类别数据生成日志的分类规则，在生成分类规则过程中，不需要人工分析日志，以及定义分类所需的正则表达式，简化了分类规则的生成流程，以及提高了分类规则的生成效率。

发明的有益效果

对附图的简要说明

附图说明

- [0024] 图1是本申请日志分类规则的生成方法第一实施例的流程示意图；
- [0025] 图2是本申请日志分类规则的生成方法第二实施例的流程示意图；
- [0026] 图3是本申请日志分类规则的生成方法第四实施例的流程示意图；
- [0027] 图4是本申请日志分类规则的生成装置较佳实施例的功能示意图模块图；
- [0028] 图5是本申请实施例方案涉及的硬件运行环境的结构示意图。
- [0029] 本申请目的的实现、功能特点及优点将结合实施例，参照附图做进一步说明。

发明实施例

本发明的实施方式

- [0030] 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。
- [0031] 本申请提供一种日志分类规则的生成方法，参照图1，图1为本申请日志分类规则的生成方法第一实施例的流程示意图。
- [0032] 本申请实施例提供了日志分类规则的生成方法的实施例，需要说明的是，虽然在流程图中示出了逻辑顺序，但是在某些情况下，可以以不同于此处的顺序执行所示出或描述的步骤。
- [0033] 日志分类规则的生成方法应用于日志分类规则的生成设备中，生成设备可以包括诸如平板电脑、笔记本电脑、掌上电脑、个人数字助理（Personal Digital Assistant，PDA）等移动终端，以及诸如数字TV、台式计算机等固定终端。在日志分类规则的生成方法的各个实施例中，为了便于描述，省略执行主体进行阐述各个实施例。
- [0034] 日志分类规则的生成方法包括：
- [0035] 步骤S10，当获取到待分类日志，确定所述待分类日志对应的日志类别，并确定各个日志类别对应的日志集合。
- [0036] 在本申请实施例中，可从网络日志服务器获取待分类日志。在网络日志服务器中，存储有交换机、路由器和防火墙等网络设备运行时产生的日志。从网络日志服务器中获取待分类日志的方法包括但不限于全量同步方法、增量同步方法和条件同步方法。全量同步方法为获取网络日志服务器存储的所有日志作为待

分类日志；增量同步方法为从网络日志服务器中获取新存储的日志作为待分类日志，此时，待分类日志为未获取过的日志；条件同步方法为可从网络日志服务器中获取特定时间段存储的日志作为待分类日志，获取其中一个或者多个设备产生的日志作为待分类日志，或者获取其中一个或者多个设备型号对应设备产生的日志作为待分类日志。可以理解的是，一个设备只对应一个设备型号，但一个设备型号可对应多个设备。

[0037] 进一步地，在获取待分类过程中，为了避免影响网络日志服务器的正常运行，在从网络日志服务器中获取待分类日志时，可根据待分类日志的文件大小、文件数量和网络日志服务器所能用的带宽限制从网络日志服务器中获取待分类日志的速率，其中，文件大小为一个待分类日志的大小。具体地，可根据文件大小和文件数量确定在当前带宽下，每次可从网络日志服务器中获取多少待分类日志，才可以在不影响网络日志服务器正常运行的情况下，快速从网络日志服务器中获取待分类日志。

[0038] 当从网络日志服务器中获取到待分类日志后，确定待分类日志对应的日志类别，并确定各个日志类别对应的日志集合。

[0039] 进一步地，步骤S10包括：

[0040] 步骤a，当获取到待分类日志后，将所述待分类日志的每一行日志中的一个信息单元确定为一个元素，以得到所述待分类日志中每一行日志对应的元素数组。

[0041] 步骤b，在所述元素数组中剥离出时间戳元素、设备名称元素和日志类别标识元素后，得到所述待分类日志中每一行日志对应的日志数组。

[0042] 具体地，当获取到待分类日志后，确定待分类日志中最小信息单元构成的日志行，在本实施例中，以英文单词作为最小信息单元，将待分类日志的每一行日志中的一个信息单元确定为一个元素，每一行日志中的全部元素对应组成了该行日志对应的元素数组。元素数组是将每行日志按照自定义的分隔符（一般为空格符），切分出单词作为数组元素，按顺序将元素存入数组构成元素数组。即采用分隔符，按照最小信息单元切分待分类日志中的每一行日志，得到每一行日志对应的元素，以得到每一行日志对应的元素数组。在元素数组中剥离出

时间戳元素、设备名称元素和日志类别标识元素后，得到待分类日志中每一行日志对应的日志数组。可以理解的是，日志数组中是不存在时间戳、设备名称和日志类别标识的。分隔符可根据具体需要而设置，如可将分隔符设置为空格符，或者设置为“|”等。时间戳是生成日志的生成时间的时间戳，设备名称是产生待分类日志对应的设备的名称，日志类别标识是产生待分类日志对应的设备对应的日志类别标识，日志类别标识是设备在出厂时就设定好的，每个设备都存在自己的日志类别标识。

[0043] 如若某一行原始日志为：“2019 Mar 22 15:50:13.196
SZ-XL-0107-G05-ISR4451-IE-AR-01 %LINK-3-UPDOWN: Interface
GigabitEthernet0/0/1, change state to up”，对应的元素数组为：[(2019),
(Mar), (22),
(15:50:13.196), (SZ-XL-0107-G05-ISR4451-IE-AR-01), (%LINK-3-UPDOWN:),
(Interface), (GigabitEthernet0/0/1,), (change), (state), (to), (up
)], 其中，[]表示数组，()表示元素，其中(%LINK-3-UPDOWN)是日志类别标识元素，(2019), (Mar), (22),
(15:50:13.196)都是时间戳元素，(SZ-XL-0107-G05-ISR4451-IE-AR-01)是设备名称元素。此时，对应日志数组为[(Interface), (GigabitEthernet0/0/1,),
(change), (state), (to), (up)]。

[0044] 步骤c，将所述设备名称和所述日志类别标识相同的待分类日志确定为属于同一日志类别的日志。

[0045] 步骤d，在同一日志类别的日志对应的日志数组中，将数组长度相同的日志数组组合成日志集合，以得到各个日志类别对应的日志集合。

[0046] 当得到待分类日志对应的所有日志数组后，将设备名称和日志类别标识相同的待分类日志确定属于同一日志类别的日志，并在同一日志类别的日志对应的日志数组中，将数组长度相同的日志数组组合成日志集合，以得到各个日志类别对应的日志集合。其中，数组长度相同的日志数组中的元素个数是相同。需要说明的是，为了减少日志集合中日志数组的数量，提高得到特征数组的效率，需要对日志集合进行去重操作，删除日志集合中多余的，完全相同的日志数组

，使日志集合中，每种日志数组都只存在一个。

[0047] 如日志类别标识为%LINK-3-UPDOWN，设备名称为SZ-XL-0107-G05-ISR4451-IE-AR-01对应的日志集合可为：[(Interface), (GigabitEthernet0/0/1,), (change), (state), (to), (down)], [(Interface), (GigabitEthernet0/0/1,), (change), (state), (to), (up)], [(Interface), (GigabitEthernet1/0/1,), (change), (state), (to), (down)], [(Interface), (GigabitEthernet1/0/1,), (change), (state), (to), (up)]。

[0048] 步骤S20，在所述日志集合中读取一个日志数组确定为目标数组，并计算所述目标数组与所述日志集合中的其它日志数组之间的第一差值。

[0049] 步骤S30，根据所述第一差值确定所述日志集合对应的第一信息模板，并检测是否只存在一种所述第一信息模板。

[0050] 当得到待分类日志中各个日志类别对应的日志集合后，在日志集合中读取一个日志数组确定为目标数组，并计算该目标数组与日志集合中其它日志数组之间的第一差值。需要说明的是，目标数组可为日志集合中的第一个数组，也可为日志集合中的任意一个数组。在计算目标数组与日志集合中其它日志数组之间的第一差值过程中，是比较日志数组与日志集合中其它日志数组相同位置的元素是否相同，如果相同位置的元素相同，则对应位置计算所得的结果为0，如果相同位置的元素不相同，则对应位置计算所得的结果为1。

[0051] 当计算得到目标数组与日志集合中的其它日志数组之间的第一差值后，根据第一差值确定日志集合对应的第一信息模板，并检测是否只存在一种第一信息模板。在本实施例中，模版是一种特殊数组，数组元素取值为0或1，信息模板用于提取同类日志事件里变化的信息，是半结构化处理的关键步骤之一，信息模板与特征模板互补，即将信息模板的0和1元素互换就成了对应的特征模板，通过求差可获得信息模板。半结构化数据是结构化数据的一种形式，它并不符合关系型数据库或其他数据表的形式关联起来的数据模型结构，但包含相关标记，用来分隔语义元素以及对记录和字段进行分层，因此，它也被称为自描述的结构。半结构化数据属于同一类实体可以有不同的属性，即使他们被组合在一起，这些属性的顺序并不重要。

- [0052] 若[(Interface), (GigabitEthernet0/0/1,), (change), (state), (to), (down)]为目标数组, 则计算目标数组与[(Interface), (GigabitEthernet1/0/1,), (change), (state), (to), (up)]之间的差值, 所得的第一差值分别为0, 1, 0, 0, 0和1, 此时, 所得的第一信息模板为[0, 1, 0, 0, 0, 1]。
- [0053] 步骤S40, 若检测到只存在一种所述第一信息模板, 则根据所述第一信息模板确定所述日志集合对应的特征数组, 并根据所述特征数组和所述日志类别对应的类别数据生成日志的分类规则。
- [0054] 若检测到该日志集合只存在一种第一信息模板, 则根据第一信息模板确定日志集合对应的特征模板, 根据该特征模板确定该日志集合对应的特征数组, 并根据该特征数组和日志类别对应的类别数据生成日志的分类规则。其中, 类别数据包括但不限于设备名称和日志类别标识。将特征数组和日志类别对应的类别数据关联存储, 即生成了日志的分类规则, 即在分类规则中, 类别数据是和特征数组关联存储的, 一组类别数据可对应一个或者多个特征数组。在所生成的分类规则中, 首先根据类别数据可确定日志所属的类别, 然后根据该类别对应的特征数组可为该类别对应的日志进行半结构化处理。
- [0055] 需要说明的是, 特征模板是同日志类别的日志数组, 元素数量相同且特定位置元素一样, 将这些特定位置的元素换成1, 其余位置换成0就构成了特征模板。特征模板是将信息模板中不含信息的元素转换成1元素, 即将信息模板中0元素对应转换成1元素, 将含有信息的1元素转换成0元素。特征数组是一种特殊数组, 数组元素取值为0或单词, 是特征模板提取日志数组后产生的结果, 反映了日志的单词分布特征。如第一信息模板[0, 1, 0, 0, 0, 1]对应的特征模板为: [1, 0, 1, 1, 1, 0], 对应的特征数组为[(Interface), 0, (changed), (state), (to), 0]。
- [0056] 本实施例通过在获取到待分类日志后, 确定待分类日志中各个日志类别对应的日志集合, 在日志集合中读取一个日志数组确定为目标数组, 根据目标数组与日志集合中的其它日志数组之间的第一差值确定日志集合对应的第一信息模板, 若检测到只存在一种第一信息模板, 则根据第一信息模板确定日志集合对应的特征数组, 根据特征数组和日志类别对应的类别数据生成日志的分类规则。

本实施例基于日志数组和模板的概念，得出特征数组，根据特征数组和日志类别生成日志的分类规则，通过模板和数组的直观性和可调节性便于用户快速调整分类规则，且不再需要生成正则表达式规则对日志进行分类，而是通过数组和类别数据生成日志的分类规则，在生成分类规则过程中，不需要人工分析日志，以及定义分类所需的正则表达式，简化了分类规则的生成流程，以及提高了分类规则的生成效率。

[0057] 需要说明的是，现有的通过正则表达式规则对日志分类过程中，不能对日志的数据进行处理，即日志中的数据还是非结构化数据，降低了用户通过日志进行数据分析和数据统计的效率。进一步地，步骤S40包括：

[0058] 步骤e，若检测到只存在一种所述第一信息模板，则确定所述第一信息模板对应的特征模板。

[0059] 步骤f，根据所述特征模板生成所述日志集合对应的特征数组，为所述特征数组中的信息元素添加对应的信息标签，根据添加信息标签后的所述特征数组和所述日志类别对应的类别数据生成日志的分类规则。

[0060] 进一步地，若检测到只存在一种第一信息模板，则确定第一信息模板对应的特征模板，根据特征模板生成日志集合对应的特征数组，为特征数组中的信息元素添加对应的信息标签，并根据添加信息标签后的特征数组和日志类别对应的类别数据生成日志的分类规则。需要说明的是，信息元素为特征数组中含有信息的元素，在特征数组中，将数值为0 的元素记为信息元素，所添加的信息标签由该信息元素表示的意思来确定。如当该信息元素表示的是IP地址时，可为该信息元素添加地址标签。在本实施例中，采用XML（Extensible Markup Language，可扩展标记语言）格式存储特征数组。

[0061] 进一步地，为了便于用户查看日志分类规则的生成过程，在生成日志的分类规则过程中，可存储该生成日志分类规则过程中产生的日志。进一步地，为了便于用户查看所生成的日志分类规则，可输出该分类规则供用户查看，用户也可根据具体需要调整该分了规则。

[0062] 通过为特征数组中携带信息的信息元素添加对应的标签，实现了将非结构化数据转换成半结构化数据，以通过半结构化数据提高了数据统计、数据分析和数

据异常跟进的效率。非结构化数据是数据结构不规则或不完整，没有预定义的数据模型，不方便用数据库二维逻辑表来表现的数据，比如各类文本。

[0063] 需要说明的是，该生成设备可为金融机构设备，金融机构设备可通过本实施例生成的日志分类规则，对金融机构设备数据处理过程中生成的日志进行分类处理。

[0064] 进一步地，提出本申请日志分类规则的生成方法第二实施例。

[0065] 所述日志分类规则的生成方法第二实施例与所述日志分类规则的生成方法第一实施例的区别在于，参照图2，日志分类规则的生成方法还包括：

[0066] 步骤S50，若检测到存在至少两种所述第一信息模板，则依次读取所述日志集合中除所述目标数组外的其它日志数组作为待计算数组。

[0067] 若检测到存在至少两种第一信息模板，则依次读取日志集合中除目标数组外的其它数据作为待计算数组。需要说明的是，本实施例中依次读取日志集合中除目标数组外的其它日志数组作为待计算数组过程中，不限于从在日志集合中从前到后开始读，还是从日志集合中从后到前开始读，只要将日志集合中除目标数组外的其它日志数组都读取到即可。

[0068] 步骤S60，计算所述待计算数组与所述日志集合中除所述待计算数组外的其它日志数组之间的第二差值。

[0069] 步骤S70，根据所述第二差值确定所述待计算数组对应的第二信息模板，并计算所述第二信息模板的出现次数。

[0070] 当在日志集合中读取到待计算数组后，计算待计算数组与日志集合中除待计算数组外的其它日志数组之间的第二差值，需要说明的是，计算第二差值的过程与计算第一差值的过程类似，在此不再重复赘述。当计算出第二差值后，根据第二差值确定待计算数组对应的第二信息模板，以及计算第二信息模板的出现次数。需要说明的是，当第二差值对应至少两个第二信息模板时，出现次数是出现次数最多的信息模板对应的次数。如第二差值对应的第二信息模板有两个，分别为A信息模板和B信息模板，而A信息模板的出现次数为2，B信息模板的出现次数为3，则本实施例中所需用到的，与第一信息模板的出现次数进行比较的是B信息模板的出现次数。

- [0071] 步骤S80, 若所述第二信息模板的出现次数大于所述第一信息模板的出现次数, 则检测所述待计算数组是否为所述日志集合中的最后一个数组。
- [0072] 当计算得到第二信息模板的出现次数后, 判断第二信息模板的出现次数是否大于第一信息模板的出现次数。若第二信息模板的出现次数大于第一信息模板的出现次数, 则存储第二信息模板, 并检测待计算数组是否为日志集合中被读取的最后一个数组。进一步地, 若第二信息模板的出现次数小于或者等于第一信息模板的出现次数, 则检测待计算数组是否为日志集合中被读取的最后一个数组。
- [0073] 步骤S90, 若所述待计算数组为所述日志集合中的最后一个数组, 则根据所述第二信息模板确定所述日志集合对应的特征数组, 并根据所述特征数组和所述日志类别对应的类别数据生成日志的分类规则。
- [0074] 若检测到待计算数组为日志集合中被读取的最后一个数组, 则根据第二信息模板确定日志集合对应的特征数组, 并根据所确定的特征数组和日志类别对应的类别数据生成日志的分类规则。需要说明的是, 在根据第二信息模板确定日志集合对应的特征数组的过程和根据第一信息模板确定日志集合对应的特征数组的过程类似, 在此不再详细赘述。
- [0075] 进一步地, 日志分类规则的生成方法还包括:
- [0076] 步骤g, 若检测到所述待计算数组不是所述日志集合中的最后一个数组, 则在所述日志集合中读取一个新的日志数组确定为目标数组, 并执行所述计算所述目标数组与所述日志集合中的其它日志数组之间的第一差值的步骤。
- [0077] 若检测到待计算数组不是日志集合中的最后一个数组, 则在日志集合中读取一个新的日志数组确定为目标数组, 并计算该目标数组与日志集合中其它的日志数组之间的第一差值。需要说明的是, 日志集合中的日志数组至少为两个, 否则无法通过计算差值得到信息模板。
- [0078] 本实施例通过在检测到存在至少两种第一信息模板后, 依次读取日志集合中除目标数组外的其它日志数组作为待计算数组, 根据待计算数组与日志集合中除待计算数组外的其它日志数组之间的第二差值确定待计算数组对应的第二信息模板, 若第二信息模板的出现次数大于第一信息模板的出现次数, 且待计算数

组为日志集合中的最后一个数组，则根据第二信息模板确定日志集合对应的特征数组，并根据特征数组和日志类别对应的类别数据生成日志的分类规则；若检测到待计算数组不是日志集合中的最后一个数组，则在日志集合中读取一个新的日志数组确定为目标数组，并重新执行计算目标数组与日志集合中的其它日志数组之间的第一差值，使最终所得的日志分类规则为所得待分类日志中出现概率最高的日志分类规则，提高了所得的日志分类规则分类日志的准确率。

[0079] 进一步地，提出本申请日志分类规则的生成方法第三实施例。

[0080] 所述日志分类规则的生成方法第三实施例与所述日志分类规则的生成方法第一和/或第二实施例的区别在于，步骤S40还包括：

[0081] 步骤h，若检测到只存在一种所述第一信息模板，则输出提示信息提示用户确认所述第一信息模板是否正确。

[0082] 步骤i，若接收到确定所述第一信息模板正确的确认指令，则根据所述第一信息模板确定所述日志集合对应的特征数组，并根据所述特征数组和所述日志类别对应的类别数据生成日志的分类规则。

[0083] 若检测到只存在一种第一信息模板，则生成提示信息，并输出该提示信息提示用户确认第一信息模板是否正确，并检测是否接收到确定第一信息模板正确的确认指令。在本实施例中，不限制提示信息的输出方式，具体地，可通过语音或者文字等方式输出提示信息。若接收到确定第一信息模板正确的确认指令，则根据该第一信息模板确定日志集合对应的特征数组，并根据该特征数组和日志类别对应的类别数据生成日志的分类规则。进一步地，若接收到修改第一信息模板的修改指令，则根据该修改指令修改第一信息模板，得到修改后的第一信息模板，并根据修改后的第一信息模板确定日志集合对应的特征数组。进一步地，若接收到确定第一信息模板错误的确认指令，则不根据该第一信息模板确定日志集合对应的特征数组，此时，可确定日志的分类规则生成失败。

[0084] 可以理解的是，在根据第二信息模板确定日志集合对应的特征数组之前，也可以输出提示信息提示用户确认第二信息模板是否正确。在输出提示信息提示用户确认第二信息模板是否正确之后的操作过程与输出提示信息用户确认第一信息模板是否正确之后的操作过程类似，在此不再详细赘述。

- [0085] 本实施例通过在得到第一信息模板后，输出提示信息提示用户确认第一信息模板是否正确，只有在接收到确定第一信息模板正确的确认指令后，才根据第一信息模板确定日志集合对应的特征数组，通过该特征数组生成日志的分类规则，进一步提高了所生成的分类规则对日志数据进行半结构处理的准确率，以及提高了所生成的分类规则对日志分类的准确率。
- [0086] 进一步地，提出本申请日志分类规则的生成方法第四实施例。
- [0087] 所述日志分类规则的生成方法第四实施例与所述日志分类规则的生成方法第一、第二和/或第三实施例的区别在于，参照图3，日志分类规则的生成方法还包括：
- [0088] 步骤S110，当获取到新日志后，提取所述新日志中的类别数据，根据所述类别数据确定所述新日志对应的分类规则。
- [0089] 在生成日志分了规则后，检测是否在网络日志服务器中获取到新日志。若在网络日志服务器中获取到新日志，则提取新日志中的类别数据，并根据该类别数据确定新日志对应的分类规则。需要说明的是，由于在分类规则中，是存在类别数据的，因此，将新日志中提取的类别数据与分类规则中的类别数据进行对比，即可确定新日志对应的分类规则。
- [0090] 步骤S120，根据所述新日志对应的分类规则确定所述新日志对应的特征数组，根据所述新日志对应的特征数组处理所述新日志，以对所述新日志进行半结构化处理。
- [0091] 当确定新日志对应的分类规则后，根据新日志对应的分类规则确定新日志对应的特征数组。其中，一个分类规则可对应一个或者多个特征数组，因此，在根据新日志对应的分类规则确定新日志对应的特征数组过程中，需要提取新日志中的日志数组，将日志数组与特征数组进行对比，若新日志的日志数组与某个特征数组中除0元素外的其它元素都相同，则可确定该特征数组为与新日志的日志数组匹配的特征数组。可以理解的是，与新日志的日志数组匹配的特征数组为与新日志对应的特征数组。在一个新日志中，会存在多行日志，因此一个新日志可能会对应多个日志数组，即一个新日志可能对应多个特征数组。需要说明的是，确定新日志对应的日志数组过程和第一实施例中确定待分类日志中

每一行日志对应的日志数组的过程相同，在此不再详细赘述。

[0092] 当确定新日志对应的特征数组后，根据新日志对应的特征数组处理新日志，即新日志对应的特征数组为新日志的日志数组中的信息元素添加对应的信息标签，以对新日志进行半结构化处理。可以理解的是，当对新日志进行半结构化处理后，由于新日志中含有特定信息的元素已被添加对应的信息标签，因此后续在对该新日志进行数据分析，可直接提取该信息标签对应的日志数据，以实现对新日志数据的快速分析。

[0093] 进一步地，若新日志无法匹配分类规则，或者匹配分类规则后无法匹配出对应的特征数组，则存储该新日志，并输出通知消息通知对应的用户新日志无法分类，或者根据该新日志重新执行步骤S10至步骤S90，以训练出该新日志对应的分类规则。

[0094] 本实施例通过在获取到新日志后，确定该新日志对应的分类规则，然后根据该分类规则确定新日志对应的特征数组，根据新日志对应的特征数组处理该新日志，对该新日志进行半结构化处理，以在对日志分类过程中对日志数据进行半结构处理，提高了对日志数据分析的分析效率。

[0095] 此外，参照图4，本申请还提供一种日志分类规则的生成装置，所述日志分类规则的生成装置包括：

[0096] 确定模块10，用于当获取到待分类日志，确定所述待分类日志对应的日志类别，并确定各个日志类别对应的日志集合；

[0097] 读取模块20，用于在所述日志集合中读取一个日志数组确定为目标数组；

[0098] 计算模块30，用于计算所述目标数组与所述日志集合中的其它日志数组之间的第一差值；

[0099] 所述确定模块10还用于根据所述第一差值确定所述日志集合对应的第一信息模板；

[0100] 检测模块40，用于检测是否只存在一种所述第一信息模板；

[0101] 所述确定模块10还用于若检测到只存在一种所述第一信息模板，则根据所述第一信息模板确定所述日志集合对应的特征数组；

[0102] 生成模块50，用于根据所述特征数组和所述日志类别对应的类别数据生成日志

的分类规则。

[0103] 进一步地，所述读取模块20还用于若检测到存在至少两种所述第一信息模板，则依次读取所述日志集合中除所述目标数组外的其它日志数组作为待计算数组

；

[0104] 所述计算模块30还用于计算所述待计算数组与所述日志集合中除所述待计算数组外的其它日志数组之间的第二差值；

[0105] 所述确定模块10还用于根据所述第二差值确定所述待计算数组对应的第二信息模板；

[0106] 所述计算模块30还用于计算所述第二信息模板的出现次数；

[0107] 所述检测模块40还用于若所述第二信息模板的出现次数大于所述第一信息模板的出现次数，则检测所述待计算数组是否为所述日志集合中的最后一个数组；

[0108] 所述确定模块10还用于若所述待计算数组为所述日志集合中的最后一个数组，则根据所述第二信息模板确定所述日志集合对应的特征数组。

[0109] 进一步地，所述读取模块20还用于若检测到所述待计算数组不是所述日志集合中的最后一个数组，则在所述日志集合中读取一个新的日志数组确定为目标数组；

[0110] 所述日志分类规则的生成装置包括：

[0111] 执行模块，用于执行所述计算所述目标数组与所述日志集合中的其它日志数组之间的第一差值的步骤。

[0112] 进一步地，所述确定模块10包括：

[0113] 第一确定单元，用于若检测到只存在一种所述第一信息模板，则确定所述第一信息模板对应的特征模板；

[0114] 生成单元，用于根据所述特征模板生成所述日志集合对应的特征数组；

[0115] 添加单元，用于为所述特征数组中的信息元素添加对应的信息标签；

[0116] 所述生成模块50还用于根据添加信息标签后的所述特征数组和所述日志类别对应的类别数据生成日志的分类规则。

[0117] 进一步地，所述确定模块10还包括：

[0118] 输出单元，用于若检测到只存在一种所述第一信息模板，则输出提示信息提示

用户确认所述第一信息模板是否正确；

[0119] 第二确定单元，用于若接收到确定所述第一信息模板正确的确认指令，则根据所述第一信息模板确定所述日志集合对应的特征数组；

[0120] 所述生成模块50还用于根据所述特征数组和所述日志类别对应的类别数据生成日志的分类规则。

[0121] 进一步地，所述确定模块10包括：

[0122] 第三确定单元，用于当获取到待分类日志后，将所述待分类日志的每一行日志中的一个信息单元确定为一个元素，以得到所述待分类日志中每一行日志对应的元素数组；

[0123] 处理单元，用于在所述元素数组中剥离出时间戳元素、设备名称元素和日志类别标识元素后，得到所述待分类日志中每一行日志对应的日志数组；

[0124] 所述第三确定单元还用于将所述设备名称和所述日志类别标识相同的待分类日志确定为属于同一日志类别的日志；

[0125] 所述第三确定单元还用于在同一日志类别的日志对应的日志数组中，将数组长度相同的日志数组组合成日志集合，以得到各个日志类别对应的日志集合。

[0126] 进一步地，所述日志分类规则的生成装置还包括：

[0127] 提取模块，用于当获取到新日志后，提取所述新日志中的类别数据；

[0128] 所述确定模块10还用于根据所述类别数据确定所述新日志对应的分类规则；根据所述新日志对应的分类规则确定所述新日志对应的特征数组；

[0129] 所述日志分类规则的生成装置还包括：

[0130] 处理模块，用于根据所述新日志对应的特征数组处理所述新日志，以对所述新日志进行半结构化处理。

[0131] 需要说明的是，日志分类规则的生成装置的各个实施例与上述日志分类规则的生成方法的各实施例基本相同，在此不再详细赘述。

[0132] 此外，本申请还提供一种日志分类规则的生成设备。如图5所示，图5是本申请实施例方案涉及的硬件运行环境的结构示意图。

[0133] 需要说明的是，图5即可为日志分类规则的生成设备的硬件运行环境的结构示意图。本申请实施例日志分类规则的生成设备可以是PC，便携计算机等终端设

备。

[0134] 如图5所示，该日志分类规则的生成设备可以包括：处理器1001，例如CPU，存储器1005，用户接口1003，网络接口1004，通信总线1002。其中，通信总线1002用于实现这些组件之间的连接通信。用户接口1003可以包括显示屏（Display）、输入单元比如键盘（Keyboard），可选用户接口1003还可以包括标准的有线接口、无线接口。网络接口1004可选的可以包括标准的有线接口、无线接口（如WI-FI接口）。存储器1005可以是高速RAM存储器，也可以是稳定的存储器（non-volatile memory），例如磁盘存储器。存储器1005可选的还可以是独立于前述处理器1001的存储装置。

[0135] 可选地，日志分类规则的生成设备还可以包括摄像头、RF（Radio Frequency，射频）电路，传感器、音频电路、WiFi模块等等。

[0136] 本领域技术人员可以理解，图5中示出的日志分类规则的生成设备结构并不构成对日志分类规则的生成设备的限定，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。

[0137] 如图5所示，作为一种计算机存储介质的存储器1005中可以包括操作系统、网络通信模块、用户接口模块以及日志分类规则的生成程序。其中，操作系统是管理和控制日志分类规则的生成设备硬件和软件资源的程序，支持日志分类规则的生成程序以及其它软件或程序的运行。

[0138] 在图5所示的日志分类规则的生成设备中，用户接口1003主要用于网络日志服务器，与网络日志服务器进行数据通信；网络接口1004主要用于后台服务器，与后台服务器进行数据通信；处理器1001可以用于调用存储器1005中存储的日志分类规则的生成程序，并执行如上所述的日志分类规则的生成方法的步骤。

[0139] 本申请日志分类规则的生成设备具体实施方式与上述日志分类规则的生成方法各实施例基本相同，在此不再赘述。

[0140] 此外，本申请实施例还提出一种计算机可读存储介质，所述计算机可读存储介质上存储有日志分类规则的生成程序，所述日志分类规则的生成程序被处理器执行时实现如上所述的日志分类规则的生成方法的步骤。

[0141] 本申请计算机可读存储介质具体实施方式与上述日志分类规则的生成方法各实

施例基本相同，在此不再赘述。

[0142] 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、物品或者装置不仅包括那些要素，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、物品或者装置所固有的要素。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、物品或者装置中还存在另外的相同要素。

[0143] 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。

[0144] 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ROM/RAM、磁碟、光盘）中，包括若干指令用以使得一台终端设备（可以是手机，计算机，服务器，空调器，或者网络设备等等）执行本申请各个实施例所述的方法。

[0145] 以上仅为本申请的优选实施例，并非因此限制本申请的专利范围，凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换，或直接或间接运用在其他相关的技术领域，均同理包括在本申请的专利保护范围内。

权利要求书

- [权利要求 1] 一种日志分类规则的生成方法，其中，所述日志分类规则的生成方法包括以下步骤：
- 当获取到待分类日志，确定所述待分类日志对应的日志类别，并确定各个日志类别对应的日志集合；
- 在所述日志集合中读取一个日志数组确定为目标数组，并计算所述目标数组与所述日志集合中的其它日志数组之间的第一差值；
- 根据所述第一差值确定所述日志集合对应的第一信息模板，并检测是否只存在一种所述第一信息模板；
- 若检测到只存在一种所述第一信息模板，则根据所述第一信息模板确定所述日志集合对应的特征数组，并根据所述特征数组和所述日志类别对应的类别数据生成日志的分类规则。
- [权利要求 2] 如权利要求1所述的日志分类规则的生成方法，其中，所述根据所述第一差值确定所述日志集合对应的第一信息模板，并检测是否只存在一种所述第一信息模板的步骤之后，还包括：
- 若检测到存在至少两种所述第一信息模板，则依次读取所述日志集合中除所述目标数组外的其它日志数组作为待计算数组；
- 计算所述待计算数组与所述日志集合中除所述待计算数组外的其它日志数组之间的第二差值；
- 根据所述第二差值确定所述待计算数组对应的第二信息模板，并计算所述第二信息模板的出现次数；
- 若所述第二信息模板的出现次数大于所述第一信息模板的出现次数，则检测所述待计算数组是否为所述日志集合中的最后一个数组；
- 若所述待计算数组为所述日志集合中的最后一个数组，则根据所述第二信息模板确定所述日志集合对应的特征数组，并根据所述特征数组和所述日志类别对应的类别数据生成日志的分类规则。
- [权利要求 3] 如权利要求2所述的日志分类规则的生成方法，其中，所述检测所述待计算数组是否为所述日志集合中的最后一个数组的步骤之后，还包

括：

若检测到所述待计算数组不是所述日志集合中的最后一个数组，则在所述日志集合中读取一个新的日志数组确定为目标数组，并执行所述计算所述目标数组与所述日志集合中的其它日志数组之间的第一差值的步骤。

[权利要求 4]

如权利要求1所述的日志分类规则的生成方法，其中，所述若检测到只存在一种所述第一信息模板，则根据所述第一信息模板确定所述日志集合对应的特征数组，并根据所述特征数组和所述日志类别对应的类别数据生成日志的分类规则的步骤包括：

若检测到只存在一种所述第一信息模板，则确定所述第一信息模板对应的特征模板；

根据所述特征模板生成所述日志集合对应的特征数组，为所述特征数组中的信息元素添加对应的信息标签，根据添加信息标签后的所述特征数组和所述日志类别对应的类别数据生成日志的分类规则。

[权利要求 5]

如权利要求1所述的日志分类规则的生成方法，其中，所述若检测到只存在一种所述第一信息模板，则根据所述第一信息模板确定所述日志集合对应的特征数组，并根据所述特征数组和所述日志类别对应的类别数据生成日志的分类规则的步骤包括：

若检测到只存在一种所述第一信息模板，则输出提示信息提示用户确认所述第一信息模板是否正确；

若接收到确定所述第一信息模板正确的确认指令，则根据所述第一信息模板确定所述日志集合对应的特征数组，并根据所述特征数组和所述日志类别对应的类别数据生成日志的分类规则。

[权利要求 6]

如权利要求1所述的日志分类规则的生成方法，其中，所述当获取到待分类日志，确定所述待分类日志对应的日志类别，并确定各个日志类别对应的日志集合的步骤包括：

当获取到待分类日志后，将所述待分类日志的每一行日志中的一个信息单元确定为一个元素，以得到所述待分类日志中每一行日志对应的

元素数组；

在所述元素数组中剥离出时间戳元素、设备名称元素和日志类别标识元素后，得到所述待分类日志中每一行日志对应的日志数组；

将所述设备名称和所述日志类别标识相同的待分类日志确定为属于同一日志类别的日志；

在同一日志类别的日志对应的日志数组中，将数组长度相同的日志数组组合成日志集合，以得到各个日志类别对应的日志集合。

[权利要求 7] 如权利要求1至6任一项所述的日志分类规则的生成方法，其中，所述若检测到只存在一种所述第一信息模板，则根据所述第一信息模板确定所述日志集合对应的特征数组，并根据所述特征数组和所述日志类别对应的类别数据生成日志的分类规则的步骤之后，还包括：

当获取到新日志后，提取所述新日志中的类别数据，根据所述类别数据确定所述新日志对应的分类规则；

根据所述新日志对应的分类规则确定所述新日志对应的特征数组，根据所述新日志对应的特征数组处理所述新日志，以对所述新日志进行半结构化处理。

[权利要求 8] 一种日志分类规则的生成装置，其中，所述日志分类规则的生成装置包括：

确定模块，用于当获取到待分类日志，确定所述待分类日志对应的日志类别，并确定各个日志类别对应的日志集合；

读取模块，用于在所述日志集合中读取一个日志数组确定为目标数组；

计算模块，用于计算所述目标数组与所述日志集合中的其它日志数组之间的第一差值；

所述确定模块还用于根据所述第一差值确定所述日志集合对应的第一信息模板；

检测模块，用于检测是否只存在一种所述第一信息模板；

所述确定模块还用于若检测到只存在一种所述第一信息模板，则根据

所述第一信息模板确定所述日志集合对应的特征数组；

生成模块，用于根据所述特征数组和所述日志类别对应的类别数据生成日志的分类规则。

[权利要求 9]

如权利要求8所述的日志分类规则的生成装置，其中，所述读取模块还用于若检测到存在至少两种所述第一信息模板，则依次读取所述日志集合中除所述目标数组外的其它日志数组作为待计算数组；

所述计算模块还用于计算所述待计算数组与所述日志集合中除所述待计算数组外的其它日志数组之间的第二差值；

所述确定模块还用于根据所述第二差值确定所述待计算数组对应的第二信息模板；

所述计算模块还用于计算所述第二信息模板的出现次数；

所述检测模块还用于若所述第二信息模板的出现次数大于所述第一信息模板的出现次数，则检测所述待计算数组是否为所述日志集合中的最后一个数组；

所述确定模块还用于若所述待计算数组为所述日志集合中的最后一个数组，则根据所述第二信息模板确定所述日志集合对应的特征数组。

[权利要求 10]

如权利要求8所述的日志分类规则的生成装置，其中，所述确定模块还用于若检测到只存在一种所述第一信息模板，则确定所述第一信息模板对应的特征模板；

所述生成模块还用于根据所述特征模板生成所述日志集合对应的特征数组；

所述日志分类规则的生成装置还包括：

添加模块，用于为所述特征数组中的信息元素添加对应的信息标签；

所述生成模块还用于根据添加信息标签后的所述特征数组和所述日志类别对应的类别数据生成日志的分类规则。

[权利要求 11]

一种日志分类规则的生成设备，其中，所述日志分类规则的生成设备包括存储器、处理器和存储在所述存储器上并可在所述处理器上运行的日志分类规则的生成程序，所述日志分类规则的生成程序被所述处

理器执行时实现如权利要求1至7中任一项所述的日志分类规则的生成方法的步骤。

[权利要求 12] 一种计算机可读存储介质，其中，所述计算机可读存储介质上存储有日志分类规则的生成程序，所述日志分类规则的生成程序被处理器执行时实现如权利要求1至7中任一项所述的日志分类规则的生成方法的步骤。

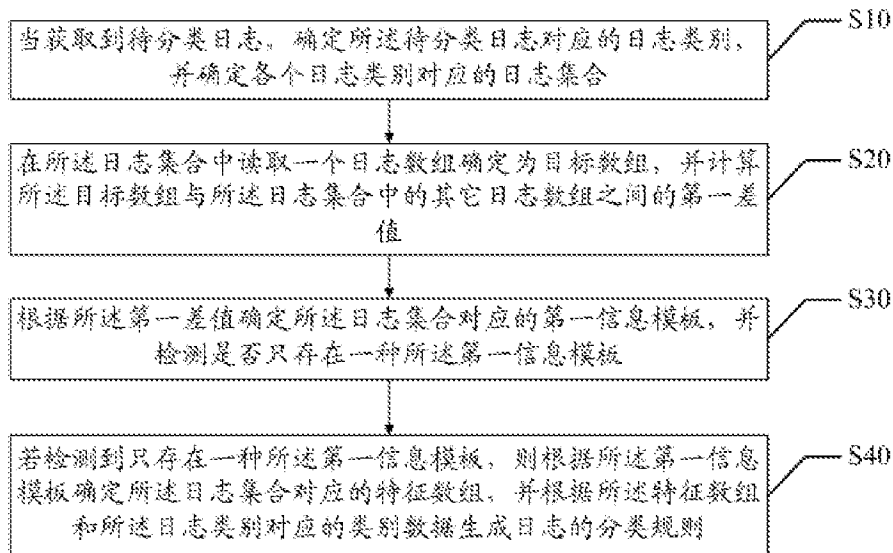


图 1

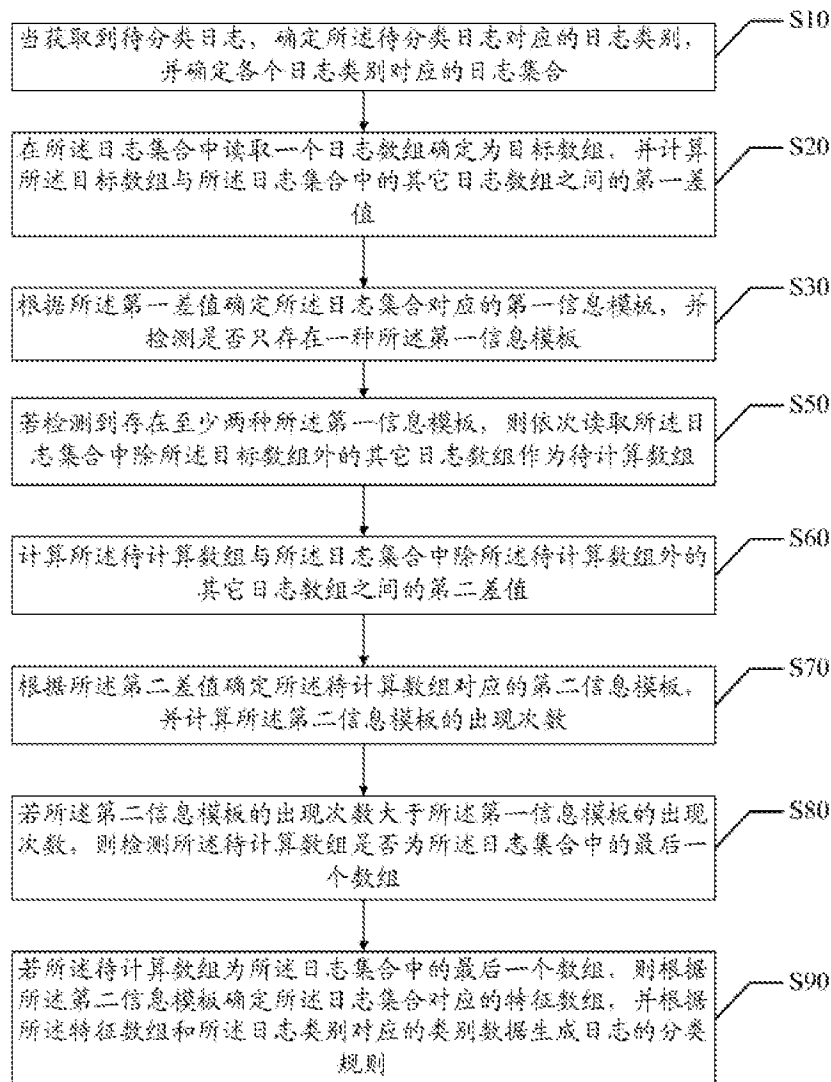


图 2

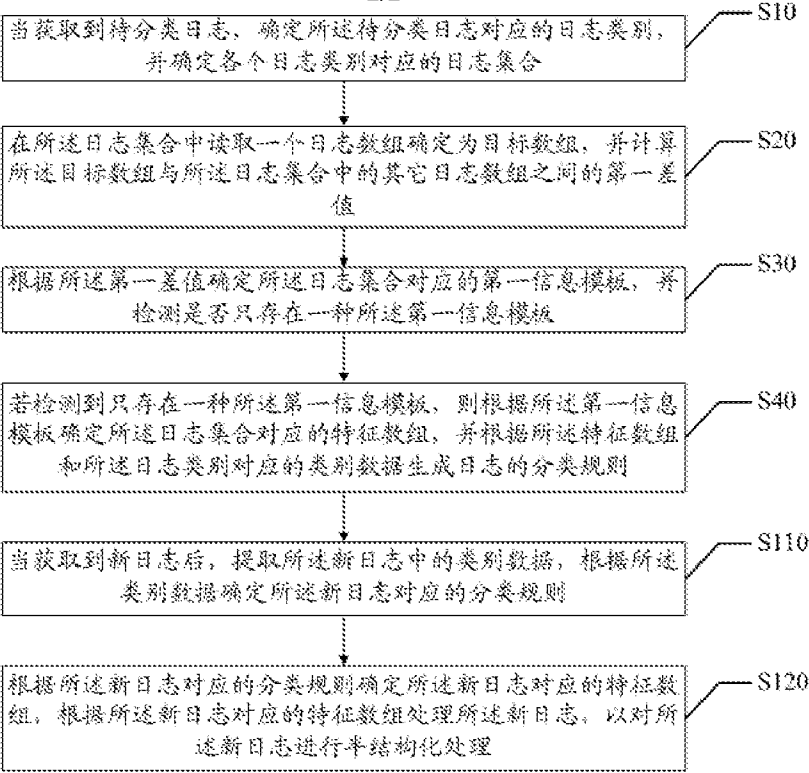


图 3

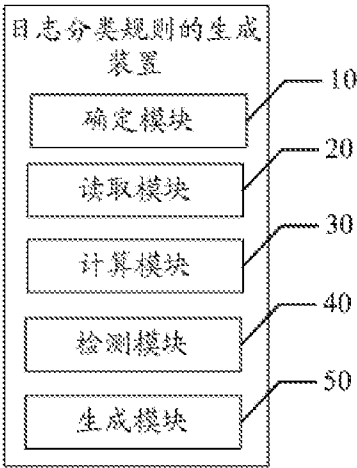


图 4

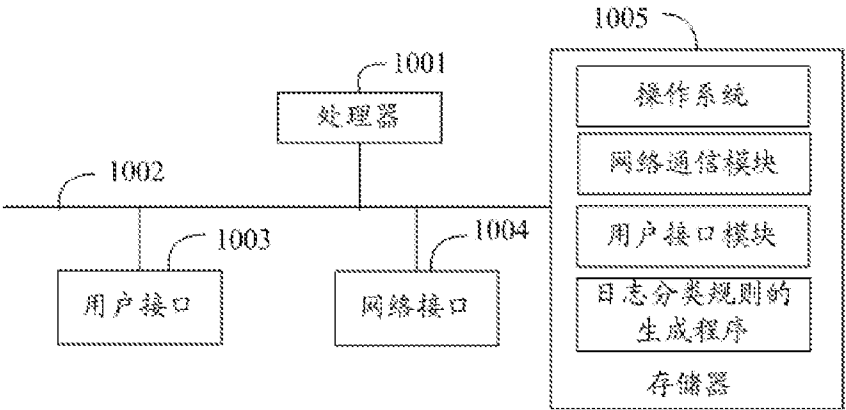


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/087835

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/17(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, CNKI, EPODOC, WPI: 日志, 文本, 分类, 规则, 模板, 正则化, 匹配, 比较, 相同, 差异, 区别, log, text, classify, rule, template, regularization, match, compare, same, difference

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110263009 A (SHENZHEN QIANHAI WEBANK CO., LTD.) 20 September 2019 (2019-09-20) claims 1-12	1-12
A	CN 107992490 A (HUAWEI TECHNOLOGIES CO., LTD.) 04 May 2018 (2018-05-04) description, paragraphs 59-129, figures 2-4	1-12
A	US 2017075932 A1 (EMC CORPORATION) 16 March 2017 (2017-03-16) entire document	1-12
A	CN 101931562 A (HANGZHOU H3C TECHNOLOGIES CO., LTD.) 29 December 2010 (2010-12-29) entire document	1-12
A	CN 105550378 A (INFORMATION CENTER, GUIZHOU POWER GRID CO., LTD.) 04 May 2016 (2016-05-04) entire document	1-12

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

15 July 2020

Date of mailing of the international search report

10 August 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/087835

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110263009	A	20 September 2019	None			
CN	107992490	A	04 May 2018	US	2019251093	A1	15 August 2019
				WO	2018076739	A1	03 May 2018
				EP	3525111	A1	14 August 2019
US	2017075932	A1	16 March 2017	CN	106547470	A	29 March 2017
CN	101931562	A	29 December 2010	None			
CN	105550378	A	04 May 2016	None			

国际检索报告

国际申请号

PCT/CN2020/087835

A. 主题的分类

G06F 16/17 (2019.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

CNPAT, CNKI, EPDOC, WPI: 日志, 文本, 分类, 规则, 模板, 正则化, 匹配, 比较, 相同, 差异, 区别, log, text, classify, rule, template, regularization, match, compare, same, difference

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 110263009 A (深圳前海微众银行股份有限公司) 2019年 9月 20日 (2019 - 09 - 20) 权利要求1-12	1-12
A	CN 107992490 A (华为技术有限公司) 2018年 5月 4日 (2018 - 05 - 04) 说明书第59-129段、附图2-4	1-12
A	US 2017075932 A1 (EMC CORPORATION) 2017年 3月 16日 (2017 - 03 - 16) 全文	1-12
A	CN 101931562 A (杭州华三通信技术有限公司) 2010年 12月 29日 (2010 - 12 - 29) 全文	1-12
A	CN 105550378 A (贵州电网有限责任公司信息中心) 2016年 5月 4日 (2016 - 05 - 04) 全文	1-12

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2020年 7月 15日

国际检索报告邮寄日期

2020年 8月 10日

ISA/CN的名称和邮寄地址

中国国家知识产权局 (ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

授权官员

董立波

电话号码 86- (10) -53961431

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/087835

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	110263009	A	2019年 9月 20日	无			
CN	107992490	A	2018年 5月 4日	US	2019251093	A1	2019年 8月 15日
				WO	2018076739	A1	2018年 5月 3日
				EP	3525111	A1	2019年 8月 14日
US	2017075932	A1	2017年 3月 16日	CN	106547470	A	2017年 3月 29日
CN	101931562	A	2010年 12月 29日	无			
CN	105550378	A	2016年 5月 4日	无			